Running Title: Dopamine consensus signaling

SYNCHRONICITY: THE ROLE OF MIDBRAIN DOPAMINE IN GENERATING WHOLE BRAIN COHERENCE, A HYPOTHESIS

Jeff A Beeler¹ and Jakob Kisbye Dreyer²

¹Queens College and the Graduate Center, City University New York, New York, NY, U.S.A.

²Department of Bioinformatics, H Lundbeck A/S, Valby, Denmark

Correspondence:
Jeff A Beeler
65-30 Kissena Blvd
Razran Rm 364
Flushing, NY 11367
jbeeler@qc.cuny.edu

KEYWORDS: phasic dopamine, tonic dopamine, dopamine architecture, reward prediction, basal ganglia, striatum, whole brain coherence, synchronous dopamine activity, consensus signal

ABSTRACT

Midbrain dopamine seems to play an outsized role in motivated behavior and learning. Widely associated with mediating reward-related behavior, decision-making and learning, dopamine continues to generate controversies in the field. While many studies and theories focus on *what* dopamine cells encode, the question of *how* the midbrain derives the information it encodes is poorly understood and comparatively less addressed. Recent anatomical studies suggest greater diversity and complexity of afferent inputs than previously appreciated, requiring rethinking of prior models. Here we elaborate a hypothesis that construes midbrain dopamine as implementing a Bayesian model selector in which individual dopamine cells sample afferent activity across distributed brain substrates, comprising evidence to be evaluated on whether current distributed neural activity is 'important.' To effectively generate a temporally resolved phasic signal, a population of dopamine cells must exhibit synchronous activity. We argue that synchronous activity across a population of dopamine cells signals *consensus*, invigorating responding to recognized opportunities and facilitating further learning. In framing our hypothesis, we shift from the question of how value is computed to the broader question of how the brain achieves coherence across distributed, parallel processing. We posit the midbrain is part of an 'axis of agency' in which the prefrontal cortex, basal ganglia and midbrain form an axis mediating control, coherence and consensus, respectively.

SIGNIFICANCE STATEMENT

Dopamine is widely associated with providing value-related signals that serve activational and teaching functions, but how these signals are derived is poorly understood. We shift focus from computing value-related signals and develop a hypothesis that suggest midbrain dopamine is sampling and integrating distributed neural activity to recognize and signal consensus. We posit each dopamine cell acts like an index of afferent activity and serves as a Bayesian unit detecting opportunity. When cells across a dopamine population converge on the same temporal patterns, dopamine cell synchrony emerges generating a consensus signal that facilitates responding to and learning about an opportunity. In this view, the fundamental role of dopamine is to signal consensus across parallel processing to facilitate unified behavioral responding.

Introduction

'Value' is a central organizing construct in adaptive behavior. Interacting with the environment, an organism needs to act on opportunities to acquire needed resources and avoid harm. While some opportunities appear immediately in the present, most opportunities lie at some point in the future and can be missed.

Consequently, a crucial component of value-based adaptive behavior is *value prediction*: from an on-going stream of sensory stimuli, determining those stimuli that indicate future opportunity for value and acting accordingly. Thus, predicting value is the cornerstone of recognizing and acting on opportunities, essentially the definition of adaptive behavior.

The midbrain dopamine system is integral to mediating value-based neural activity underlying adaptive behavior (Berridge, 2007; Salamone and Correa, 2012; Schultz, 2007; Wise, 2004). Evidence continues to accumulate that midbrain dopamine can signal (i) obtained value, (ii) value prediction and (iii) value prediction errors (Berke, 2018; Watabe-Uchida et al., 2017). Dopamine activity has been shown to be sufficient for mediating learned associations between stimuli and reward (Steinberg et al., 2013), demonstrating it can serve a crucial role as a value teaching signal (Eshel et al., 2016; Hart et al., 2014; Kishida et al., 2016; Schultz, 1998; Sharpe et al., 2017). Dopamine has been shown to modulate value-based behavioral choice and to increase (energize) responding to value predictive stimuli (da Silva et al., 2018; Hamid et al., 2015, 2015; Ko and Wanat, 2016; Roitman, 2004). Dopamine has long been believed to provide a simple, scalar signal broadcast widely across the brain (Bar-Gad et al., 2003; Joshua et al., 2009; Schultz, 1998), essentially uniform across dopamine cells (Eshel et al., 2016; Schultz, 1998). Insofar as decision-making is distributed across neural substrates, value must be mapped onto those distributed processes. By broadcasting a uniform, scalar signal widely, dopamine is widely believed to 'infuse' value information across neural substrates, mediating value-based modulation of processing and plasticity in target regions.

The nature of dopamine's contribution to value-based adaptive behavior continues to be contentious. While evidence demonstrates that dopamine can encode prediction errors (Eshel et al., 2015; Hart et al., 2014; Watabe-Uchida et al., 2017), dopamine can signal value *per se* as well, described by Hamid et al (Berke, 2018; Hamid et al., 2015) as instantaneous value signal. While dopamine can signal receipt of primary reward (obtained value), the degree to which this signal declines if reward is expected, representing a 'surprise signal'

is less clear. Dopamine signals are not restricted to reward. Bromberg-Martin has proposed that dopamine can provide alerting, reward and aversive signals (Bromberg-Martin et al., 2010), though these can be construed as value-related. Dopamine responds to novelty (Horvitz, 2000; Overton et al., 2014), though this can be construed as an ‘information bonus’ (Kakade and Dayan, 2002). Dopamine signals have been linked to motor activity independent of reward (Dodson et al., 2016; Howe and Dombek, 2016; Pasquereau and Turner, 2015). Other evidence suggests that moment-to-moment motivational states can modulate dopamine (Sato et al., 2003; Syed et al., 2016). Dopamine prediction error signals are intimately linked to *timing* of events (Langdon et al., 2018; Pasquereau and Turner, 2015; Starkweather et al., 2017; Takahashi et al., 2016) and, conversely, changes in dopamine have been demonstrated to affect timing (Soares et al., 2016). Dopamine has been demonstrated to play a role in arousal (Eban-Rothschild et al., 2016), suggesting an additional dimension in dopamine signaling. The notion that dopamine cells provide a uniform, broadly distributed signal is increasingly contested (Lammel et al., 2014; Morales and Margolis, 2017; Roeper, 2013; Sanchez-Catalan et al., 2014; Willuhn et al., 2012), with data supporting both homogeneity across dopamine cell activity (Eshel et al., 2016) and diversity in dopamine signals in target regions (Lerner et al., 2015; Matsumoto and Hikosaka, 2009; Parker et al., 2016). In computational models, dopamine signals have been construed to represent abstracted value information divorced from details of stimuli and actions, i.e., ‘model-free’ (Montague et al., 1996). It is becoming increasingly clear, however, that dopamine encodes and responds to *features* of stimuli and reward (Takahashi et al., 2017) and can be ‘model-based’ (Daw, 2012; Daw et al., 2011; Langdon et al., 2018).

Many excellent reviews discuss the various issues outlined (Berke, 2018; Schultz et al., 2017; Watabe-Uchida et al., 2017). We will focus on a comparatively less discussed question: how is the value information contained in a phasic dopamine signal derived? The outsized role ascribed to dopamine in value-based adaptive behavior provokes the question *how does dopamine ‘know’ what it knows?* Alternatively, *who teaches the teacher?*

Current ideas on derivation of dopamine value-related signals

Earlier models on how dopamine signals were derived were based on variants of the algorithmic actor-critic model (Barto, 1995) in which values associated with stimuli and actions are used to select actions (the actor, implementing a policy), which are updated following the outcome (the critic, maintaining appropriate valuations). These models presupposed dopamine to encode a prediction error and so the task was to determine how the quantities of (i) value predicted at time $t-1$, V_{t-1} , (ii) predicted future value at time t , V_t , and (iii) obtained value at time t , r_t were provided to dopamine cells to generate the error signal comprised of $r_t + V_t - V_{t-1}$. A key problem to be solved with these models was how a representation of value prediction at $t-1$ was temporally sustained and delivered at time t in order to facilitate the computation. Earlier models centered around the basal ganglia connections with the midbrain (Brown et al., 1999; Contreras-Vidal and Schultz, 1999; Houk et al., 1995). While elegant, these models did not always align with anatomy (Joel et al., 2002). In these models, the basal ganglia provided predicted values while obtained reward, r_t , was provided by excitatory drive(s), such as the lateral hypothalamus or PPTg. Subsequent models incorporated the ‘spiraling’ connectivity (Haber et al., 2000) of midbrain dopamine and striatal territories (Haruno and Kawato, 2006) and assigned a role for excitatory inputs, such as the PPTg, in value representations (Kawato and Samejima, 2007). Tan and Bullock elaborated an earlier model (Tan and Bullock, 2008), adding in a pathway from striatal matrix MSNs to dopamine cell dendrites in the SNR to account for ramping of dopamine cell activity in the intervals between cue and reward, interpreted as encoding uncertainty.

Hazy and colleagues (Hazy et al., 2010, 2007) account for patterns of DA activity with a model where the elemental components are primary and learned value (PVLV), quantities instantiated in both excitatory and inhibitory pathways. Rather than looking for neural correlates of quantities in the TD algorithm (e.g., V_t), the PVLV model was built more around neural substrates known to contribute to Pavlovian learning, such as the amygdala and ventral striatum, and abstracting what these regions were computing and how their output, funneled to the midbrain, could generate observed patterns of dopamine signaling. Continuing with a more expanded range of neural contributions to the midbrain dopamine system but returning to instantiation of TD and prediction errors, Vitay and Hamker (2014) proposed a model, like O’Reilly and colleagues, incorporating more inputs than earlier actor-critic models. In contrast to the growing complexity of these models, a recent

hypothesis suggest that GABAergic cells in the midbrain receive expected reward information that ramps up as reward approaches and drives local inhibition of dopamine cells (Eshel et al., 2015). This local inhibition derived from expected value then ‘subtracts’ from excitatory drive arising from obtained value, providing a simple subtractive mechanism for deriving an RPE. Though elegant, the model begs the question of where the expected value signal in midbrain GABA cells is derived, effectively returning us to prior models with an added layer of complexity in the midbrain.

The problem of anatomy

All the models described above are based on the notion that defined quantities are computed in different brain regions and then sent to the midbrain for dopamine computations. Given that the midbrain is believed to broadly transmit a uniform value-related signal (Bar-Gad et al., 2003; Schultz, 1998), with considerable evidence demonstrating homogeneity of dopamine cell activity (i.e., work of Schultz; Eshel et al., 2016), all of these theories run headlong into a fundamental question: how is the information/quantities putatively computed in different regions uniformly delivered across dopamine cells? This question takes two forms: (1) how is the *output* of afferent regions made uniform and/or (2) how is the *connectivity* to dopamine cells distributed uniformly. Most of the regions contributing to midbrain dopamine— just taking the amygdala as an example— exhibit ensemble encoding where patterns of particular neurons firing phasically (and others not firing) encode specific stimuli. Models that suggest region x computes and sends some quantity (e.g., V_t) to the midbrain implicitly suggest a scalar output inconsistent with ensemble encoding operating in the afferent region. Concretely, the axons leaving the amygdala and streaming to the midbrain reflect the combinatorial ensemble activity of the neurons from which they arise rather than uniformly encoding of a scalar. This means at any given moment, some amygdalar axons to the midbrain will be highly active and some will not. If afferent regions are not delivering a uniform scalar across their projections to the midbrain, but instead reflect different patterns of activation across discrete axonal inputs, then how is this *pattern* of activity distributed across dopamine cells to facilitate uniformity that allows dopamine to render computations and transmit a scalar signal? It is unlikely that every axon from region x connects with every dopamine cell to deliver uniform information. But if axons from region x are distributed, then every differential pattern of combinatorial activity

would stimulate different sets of dopamine cells, facilitating ensemble coding in dopamine cells, which is contrary to what is observed, i.e., population encoding of a uniform, scalar value. This problem is compounded when considering that midbrain dopamine receives afferents from *multiple* sources. In this case, in addition to patterned, combinatorial distribution of information from region *x*, we are faced with combinations of combinatorial distributions as, for example, the distribution pattern of region *x* and *y* inputs generate second order combinations of the two inputs. And yet, there is no evidence that midbrain dopamine operates using ensemble encoding where the activity of different combinations of dopamine cells encode and transmit discrete information. On the contrary, evidence favors population-based encoding.

Recent anatomical studies highlight this conundrum. In recent years, a spate of elegant anatomical studies have reinvestigated the connectivity of midbrain dopamine using newer tracing methods, including TRIO to map input/output relationships. In these studies, what is perhaps most striking is the extent to which the inputs to the dopamine cells, regardless of their projection targets, arise widely from across the brain (Beier et al., 2015; Lerner et al., 2015; Watabe-Uchida et al., 2012; Yetnikoff et al., 2014). The authors of these studies have focused on *biases* in inputs— for example, that the substantia nigra compacta dopamine cells projecting to the DMS and DLS exhibit a bias for inputs from those regions, respectively (Lerner et al., 2015). This interpretive focus reinforces a sense of input-output organization. If one steps back and looks at the overall pattern, however, what is striking is the diversity and apparent *non-specificity* of afferent inputs (Yetnikoff et al., 2014). For example, in the case of the Lerner et al (2015) study, the DMS-projecting dopamine cells receive more *total* inputs from the DLS than the DMS (Lerner, Fig 1). Watabe-Uchida et al (2012) note that afferents into the midbrain, rather than reflecting distinct connectivity between functional brain regions seems to arise in *bands* across the brain that do not respect functional regions, a characterization that would be more consistent with ‘random sampling’ of broad brain activity.

These problems are further highlighted in work by Tian et al (2016) in which they recorded from different afferent inputs to the midbrain, as well as midbrain dopamine cells, and correlated firing patterns in afferent projections with task activities (cue, receiving reward, etc.). Rather than any clear segregation in afferent regions of specific quantities — reward, predicted value, error signals— they found that all of these quantities were distributed across the inputs tested with many cells showing mixed patterns of activities (e.g., both cue

and reward activation), seemingly belying any notion of tidy segregation of discrete quantities being computed in localized neural substrates and sent to the midbrain. Instead, the emerging challenge seems to be, given the apparent diversity of afferent inputs and potential combinatorial distribution of these afferents across individual dopamine cells, how midbrain dopamine cells exhibit emergent, coherent population based activity. That is, in a sense, how does midbrain dopamine bring order out of apparent chaos?

The necessity of dopamine cell cooperation

The core problem, then, is how do dopamine cells take diverse, heterogeneous inputs, apparently encoding a range of information, and generate the frequently reported uniformity of activity that gives rise to a population generated scalar? We will frame this problem as the question of synchrony: that midbrain dopamine cells, as a population, must implement a process of selective integration of diverse inputs that, in turn, generates synchronous phasic dopamine cell activity; that is, in a sense dopamine cells must 'come to a consensus' in their phasic responses to input.

The midbrain dopamine nuclei contain a modest number of neurons, in rodents approximately 20,000 - 40,000 TH+ neurons in A8-10 in total, estimated to be between 400,000 - 600,000 in humans (Björklund and Dunnett, 2007). Each neuron, however, projects extensively across a large territory of target region(s). Individual SNC dopamine neurons are estimated to make 100,000 - 250,000 synapses in the rat and 1-2 million in humans (Bolam and Pissadaki, 2012), cover 6% of striatal volume and affect approximately 75,000 MSNs (Matsuda et al., 2009). VTA dopamine neurons make fewer connections, estimated between 12,000 - 30,000 synapses, with less axonal branching, though the characteristic of distributed axonal coverage is the same, if to a lesser extent.

Conversely, each MSN synapses with between 441 to 1058 dopamine terminals (Arbuthnott and Wickens, 2007; Bolam and Pissadaki, 2012) and each MSN is estimated to be under the influence of 100-200 different dopamine neurons (Matsuda et al., 2009). The diffusion distance provides a measure of how far from the release site a DA molecule can move and thus the distance at which it can activate postsynaptic receptors. The diffusion distance is measured to be around 7-10 μm (Gonon et al., 2000; Nicholson, 1985). Thus, a dopamine molecule binding to any given receptor could be released from any of several hundreds of nearby

terminals and on the order of 100 different dopamine neurons (Arbuthnott and Wickens, 2007; Cragg and Rice, 2004; Dreyer et al., 2010).

This mechanism of diffusion with overlapping spheres of influence that modulate extracellular dopamine concentration is characteristic of volume transmission attributed to dopamine (Agnati et al., 1995; Zoli et al., 1998). Rather than releasing neurotransmitter discretely in a synapse, which then transmits an ultra-targeted intra-synaptic signal quickly terminated by reuptake, dopamine operates by modulating its extracellular concentration (Cragg and Rice, 2004; Garris et al., 1994; Gonon, 1997; Moss and Bolam, 2008). The dopamine transporter determines the time-scale of clearance, playing a key role in regulating the temporal bandwidth of the DA signal. Other parameters of release, such release probability and quantal size, can also alter the temporal and spatial properties of volume transmission (Sulzer et al., 2016). Both volume transmission and the broadly distributed overlapping projections of dopamine cell axons, as well as the comparatively small numbers of neurons, suggest that to effectively transmit a signal, dopamine cells must work cooperatively.

The problem of synchrony

A dopamine signal encoded through modulation of cell spiking is decoded in target regions by modulation in receptor occupancy, determined by fluctuating extracellular dopamine concentrations: thus, encoding-decoding fundamentally reflects a transformation from spike rate (frequency modulation) to fluctuations in extracellular dopamine concentration (amplitude modulation). While many factors shape this transformation, including spike patterns, the critical determinant is the degree of synchrony among spiking neurons relative to the clearance rate of the DAT (Dreyer, 2014; Dreyer et al., 2010; Dreyer and Hounsgaard, 2013). When spike activity is correlated between neurons on a timescale similar to K_m/V_{max} (corresponding to 100 ms in nucleus accumbens and 40 ms in dorsal striatum), release will integrate and generate temporally resolved peaks and troughs in extracellular dopamine that maximize information transfer to postsynaptic targets. When spike activity is asynchronous, including asynchronous bursting activity, dopamine concentration will be distributed across time, simply reflecting the average spike rate. We will refer to 'tonic' and 'phasic' in terms of whether population spiking activity generates temporally resolved fluctuations in extracellular dopamine (phasic, in

phase) or a distributed averaging of activity (tonic, dopamine tone), regardless of whether the spiking of individual neurons occurs in singlets or bursts (Dreyer et al., 2016).

In this formulation, tonic and phasic signaling are not defined by patterns of spike activity, but by the degree of synchrony among a population of dopamine cells comprising a projection to a target region (Dreyer et al., 2016, 2010). Synchrony partitions dopamine release into temporal resolution while asynchrony distributes release across time. In this view, rather than reflecting two distinct modes of signaling, tonic and phasic represent two ends of a spectrum in temporal partitioning of dopamine cell activity.

Obligate vs. emergent dopamine cell synchrony

Dopamine signaling has been construed as arising from uniformity in bursting activity, consistent with electrophysiological evidence where dopamine cell recordings show a high percentage of tested cells exhibit similar and consistent responses to stimuli (e.g., Eshel et al., 2016), as well as with electrochemistry recordings of released dopamine, which fundamentally reflect the combined activity of hundreds of terminals. The question is not whether dopamine cells can fire synchronously to generate a signal, but *how* and *to what extent*.

We will differentiate between two modes of synchrony: obligate synchrony in which the activity of one neuron is in some way yoked to other neurons such that they effectively function as a single unit vs. emergent synchrony in which cells function autonomously, but independently come to generate the same temporal patterns of activity. While initially it was thought that dopamine cells are in some way coupled generating obligate synchrony, accumulating evidence favor emergent synchrony.

One mechanism originally proposed to account for dopamine cell synchrony was the linkage of dopamine cells by gap junctions (Grace and Bunney, 1983). These direct cell to cell connections, however, may be less prevalent than originally thought, with estimates suggesting only ~20% of dopamine cells are connected by gap junctions (Vandecasteele, 2005). Evidence suggest that these connections may provide a low-pass filter such that the activity in one neuron can influence firing rate in a connected neuron, but without transmitting yoked action potentials; that is, no spike synchrony was observed in recorded pairs and mutual information was weak (Vandecasteele, 2005). Dopamine cells connected by gap junctions connect to only 1-4 (Grace and

Bunney, 1983; Vandecasteele, 2005) cells. As a consequence, there is little to no evidence that these gap junction connections induce an obligate synchrony, though modeling suggests such gap junctions could entrain network activity (Komendantov and Canavier, 2002) and may play a role in promoting emergent synchrony.

In studies examining pairs of dopamine cells, roughly 25% of dopamine cell pairs show significant correlation in spiking rate (Hyland et al., 2002; Li et al., 2011; Morris et al., 2004; Wilson et al., 1977) with little evidence of direct spike-to-spike correlation (Morris et al., 2004). The degree of correlation can be modulated by reward associated stimuli, learning and pharmacological manipulations. Using noise correlation analysis (correlation between two cells in their trial-to-trial variation from average response), Joshua et al (2009) demonstrate that correlation in spike variability between dopamine cells increases with salient stimuli (cue, outcome). Kim et al (2012) show an increase from 34% (noise) correlated cell pairs on initial exposure to a task to 49% and 66% after 8 and 16 weeks of training, respectively. Finally, Li et al (2011) demonstrated that application of nicotine increase the percentage of correlated dopamine cell pairs from 21% to 44%. These data suggest emergent rather than obligate synchrony.

Many have previously suggested that synchronous dopamine cell bursting activity is necessary to generate temporally resolved extracellular dopamine signals in target regions (Arbuthnott and Wickens, 2007; Dreyer et al., 2010; Owesson-White et al., 2012; Venton et al., 2003). Here we will take this one step further and suggest that the selection and integration of inputs necessary for the emergence of synchrony reflects a core, fundamental function of midbrain dopamine.

Reconceptualizing midbrain dopamine as sampling distributed activity across the brain

Consider the axons streaming from each afferent region to the midbrain as a vector where each axon delivering a continuous, rate-based signal is an element in the vector and the vector represents the total output (to the midbrain) of that afferent region. We posit that these axons, each an element in the afferent vector, are distributed randomly to dopamine neurons (Fig 1A). If each dopamine neuron is, in turn, viewed as a vector composed of its various inputs, each dopamine cell represents a random sample of afferent activity where the afferent vectors from input regions collectively comprise a multidimensional space (Fig 1B). We construe this space as event driven, i.e., that neural activity is continuous and constantly responding to on-going

sensorimotor information. An event such as a cue-light, then, would be represented in this multidimensional space as a composite of the vectors representing individual afferent regions. Stimuli, actions, events, memories— everything encoded by the brain— would generate a unique composite in this multidimensional space.

The task of the midbrain, roughly speaking, is to fire in synchrony as a population when something ‘important’ appears in this multidimensional space. Each individual dopamine cell receives a randomly sampled subset of this multidimensional space and each individual cell must determine when ‘something important’ is occurring in its sample of afferent activity. In effect, each dopamine cell is a sample-based model of distributed activity in the brain (Fig 1B). When more dopamine cells detect something ‘important’ at the same time, this generates increased synchronous activity and a concomitant increase in temporally resolved extracellular dopamine corresponding to temporal occurrences of stimuli. In a sense, we are proposing that midbrain dopamine serves somewhat like an index in the stock market, where the index provides a single measure of a population of stocks, except the analogy might be better if each dopamine cell were thought of as an index and the midbrain as an index of indices, sensitive to the degree to which changes in value (input activity) occur in the same direction simultaneously. In this sense, the (rodent) midbrain with approximately 20,000 dopamine cells per side has approximately 40,000 sample-based models of ‘what is going on in the rest of the brain.’

The precise distribution of afferents in the sense described above, i.e., how the *total* output (each and every axon) from any region to the midbrain is distributed across dopamine cells, has not been characterized and is not trivial. However, as newer methods have provided more refined insight into midbrain connectivity, including input-output mappings, the extant data could be construed as consistent with an essentially random distribution, as proposed above; in fact, extant data makes it difficult to argue for *non-random* distributions, such as had been proposed in prior models. The notion that dopamine cells individually sample afferent input space— broadly construed as ‘what is happening in the rest of the brain’-- could be implemented in a variety of forms. For example, there are both direct (monosynaptic) and indirect inputs to dopamine cells. A structure such as the PPTg, which itself receives broad inputs from many regions, could serve a special role, for example as an ‘amplifier’ or as a coincidence detector (i.e., that the integrated activity the PPTg coincides with monosynaptic input into dopamine cells). When we speak of ‘dopamine cells’ as a single population, this is

Figure 1 *Random distribution of afferents to dopamine cells as sampling the input space.* (A) Projections from each afferent region construed as a vector randomly distributed to individual dopamine cells. (B) Afferent projections to midbrain comprise a total vectorized input space from which individual dopamine cells represent random samples of that space.

shorthand for the observation that dopamine cells need to fire together as a population; this does not mean that all dopamine cells in the midbrain belong to the same population. For purposes of the hypothesis being developed here, we focus on the core idea: that dopamine cells are sampling afferent activity across the brain and generate a meaningful signal as a population when individual dopamine cells converge and detect ‘importance’ at the same time, giving rise to emergent synchrony.

Information content of afferent input: a case for implicit value

What is the information content of that afferent input space being sampled by midbrain dopamine? Prior theories have construed abstracted value quantities— e.g., V_t , V_{t-1} , r_t . We will argue that *value is not abstracted and encoded as a discrete quantity but rather implicit in afferent activity*. Concretely, we argue that dopamine is receiving whatever type of information is being represented in various afferent regions, whether this be spatial information (hippocampus), stimuli-valence information (amygdala), cached, model-free information (dorsolateral striatum), sensory information, motor information, timing information, or even explicit value information *per se* (example, OFC inputs). Midbrain dopamine is, effectively, receiving a sample of distributed activity across much of the brain in all its complex, multifaceted glory. We suggest that value-related activity is apparent because value is *intrinsic* to neural processing and activity. For example, the activity in M1 in a tennis player about to make a serve is highly focused on generating overlearned movements to produce a ‘good serve.’ That selective activity reflects value because M1 has learned that this particular activity, rather than *other* activity, produces the desired results. Cells in the hippocampus that represent landmarks exhibit high activity because such landmarks have value in navigation. In short, we argue that to a large extent, neural activity is inseparable from value and, in a sense, neural activity *is* value insofar as ‘this neural activity arose and exists because it facilitates a positive outcome.’ This is an ideal, of course: actual results may vary.

The idea that ‘value’ is intrinsic to neural activity generates many questions: in any given afferent region, how is value reflected— instantaneous value, future value, value prediction errors? The study by Tian et al (Tian et al., 2016) that found that all afferent regions tested encoded a mixture of different types of value signals suggest that the characteristics of value inherent in neural activity in different regions likely varies and that how value is represented would need to be determined region by region. For example, in a spatial map in the hippocampus it may well be that a rat at point A in a maze causes activity in the cell representing a ‘future’ location in the maze, point B, providing a natural substrate for a series of states associated with V_t (Foster et al., 2000). The prefrontal cortex can sustain selected stimuli in working memory (Durstewitz et al., 2000; Fuster and Bressler, 2012; Miller and Cohen, 2001), essentially allowing value expectations to be extended in time. Ascertaining the different forms in which value is incorporated into neural activity in different afferent regions is an on-going task and beyond the scope of this hypothesis. Our key point is that rather than trying to localize

discrete, abstracted quantities such as V_t to particular afferent inputs, we suggest that the data are more consistent with the notion that value is computed across many different substrates and, as we argue, may be largely intrinsic to neural processing; that is, value is not a discrete quantity but is distributed and implicit in selective neural activity across all the different domains of representation the brain encodes. In the hypothesis we are building, then, midbrain dopamine is not taking discrete quantities and generating a value-related signal but is, rather, *integrating* implicit value information, in many forms, from multiple afferent sources. Moreover, the *content* of that value information is not abstract or divorced from the information being represented by afferent regions— space, stimuli, self, memory, expectations (and surprise) and so on.

By sampling and integrating broad activity across the brain, in which value is encoded intrinsically in various forms, the dopamine signal becomes a composite of distributed ‘value-related’ activity. The magnitude of the signal depends not only upon the firing rate of individual dopamine cells, but also upon the degree to which they exhibit synchronous activity. In essence, when a population of dopamine cells all ‘come to the same conclusion’ that ‘what is going on now is important,’ a temporally resolved phasic signal is generated in target regions. Succinctly, we propose that midbrain dopamine is integrating value information sampled from a multidimensional input space and generating a time-varying, population-based scalar signal detecting temporal convergence of distributed value activity across that input space. Concretely, if at time t the cortex, PPTg, amygdala, hippocampus and other afferent contributors all recognize value *at the same time*, this generates a phasic dopamine signal. While it appears we are arguing for a value signal, it must be noted that ‘what is important now’ may include *errors* in prediction. What we are arguing for, and will develop below, is that dopamine, rather than signaling value *per se*, is signaling *consensus* across distributed afferent regions that all intrinsically encode value-related information in multiple forms.

Searching for the ghost in the machine: where are the roots of value

Thus far, the hypothesis we have sketched out is simplistic. Simply equating of the amount of afferent activity with value is problematic. For instance, in the tennis serve example, the overall activity of M1 is likely to be *less* in a well-learned behavior, not more, as extraneous activity is inhibited and only task-relevant activity enhanced. Moreover, if activity alone intrinsically encoded value, isn’t the brain *a/ways* active? And don’t we

encode low value stimuli? More fundamentally, given that dopamine signaling is widely construed as providing the teaching signal by which other brain regions learn about value (Montague et al., 1996; Morris et al., 2004; Schultz, 2007), have we put the cart before the horse? That is, if dopamine signals consensus about value across afferent regions, then wouldn't a teaching signal be redundant at that point . . . and how did those afferent regions select intrinsically valuable activity in the first place?

Both prior theories and ours present a chicken and egg dilemma: if dopamine, as widely construed, is a primary substrate for mediating value learning but dopamine value-related teaching signals depend upon value learning already having occurred in afferent regions, how can this be? This points to a larger question: how is value determined in the brain broadly? The notion that dopamine is a 'reward system' is intended specifically as an answer to this question, but looked at closely this answer merely begs the question. In essence, the problem is locating the ghost in the machine: value-related information in various forms appears to be widely distributed throughout the brain, but what is the mechanism informing this distribution or representation of value? While an elegant case has been made for midbrain dopamine distributing value information, asking how dopamine knows what it knows leads circularly back to the same problem of distributed value: the ghost remains at large. To tackle this conundrum, we will shift our frame of reference and core question. Broadly speaking, the function associated with dopamine involves recognizing available reward opportunity, weighting decision-making to act on the opportunity, energizing responding and facilitating learning about the context and predictors in which the reward occurred: in short, to facilitate adaptive choices in the face of opportunity. It is not unreasonable to suggest that this is fundamentally what the *entire brain* evolved to do. Instead of asking how value is determined— in the midbrain or elsewhere— we will turn to another fundamental question in neuroscience: how is distributed, parallel processing brought into coherence such that information in different regions is integrated to facilitate adaptive behavior? We suggest this will lead back to an answer to the value question.

The whole brain challenge

The brain must process a continuous stream of on-going sensory information and from this parse actionable stimuli and emit advantageous responses (Cisek and Kalaska, 2010). This requires building a model of the world *in relation* to the organism. Building such a model and continuously parsing an on-going

sensory stream is complex, but must be done rapidly to ensure the animal responds *in time* (Cisek, 2007). Processing information in a distributed, parallel fashion facilitates rapid, efficient responding. In this sense, each region/substrate forms its own model of information (Fig 2). Simplistically, the hippocampus forms a spatial model: where actionable stimuli and events are located (Gauthier and Tank, 2017; Mamad et al., 2017; Stachenfeld et al., 2017). The amygdala forms a model of the valence of stimuli in relation to the organism (Beyeler et al., 2018; O'Neill et al., 2018; Pryce, 2018). The parietal cortex forms a model of ego-centric sensorimotor space (Buneo and Andersen, 2006; Cui, 2014; Takiyama, 2015; Whitlock, 2014), and so on.

A word clarifying our use of the term 'model' is in order. The simplistic examples we gave above (i.e., "hippocampus forms a spatial model") suggest that 'model' refers to encoding particular aspects of the world. While this sense of the word 'model' is true— the visual cortex does encode visual stimuli, the auditory cortex sound stimuli, motor cortex musculoskeletal state— in another sense it is too limiting. In our earlier discussion, we suggested that each dopamine cell serves as a sampled model of 'activity across the brain.' At risk of pointing out the obvious, a 'model' is a set of neurons whose time-varying activity encodes changing states in the world, the organism, or some relationship between the two. While such encoding may have obvious correlates to stimuli— a visual stimulus, position in space, position or trajectory of arm, hunger— it may also reflect less obvious representations, such as combinations of stimuli or abstractions, such as value or goals. A model may encode past, present or future expected states, or some combination. It may encode stimulus-response relationships or action-outcome relationships. The suggestion by Hamid et al (Hamid et al., 2015) that dopamine represents an on-going instantaneous value signal is the same as saying that dopamine output is a continuous model of reward availability in the environment. Sometimes what a model encodes may be difficult to define. For example, the basal ganglia have been proposed to serve as a substrate for dimension reduction (Bar-Gad et al., 2003), but precisely defining what is encoded once dimensions are compressed is far from clear.

In order to adaptively behave in the world, an organism needs an adequate internal representation of its environment, its own organismal state and knowledge of how the two interact: a model of self-in-world. A total, wholistic model of self-in-world— encompassing all possible characteristics, dimensions, relationships and potential states, internal and external— is literally an infinite, open-ended computational problem, and entirely

DISTRIBUTED MODELS OF SELF AND WORLD

Figure 2 *Distributed models of self-in-world*. Simple illustration of the idea that incoming sensory information is distributed throughout the brain with different neural regions processing that information into parallel models that encode different aspects and characteristics of self and world. Abbreviations: DLPFC, dorsolateral prefrontal cortex; OFC, orbital frontal cortex; DLS, dorsolateral striatum; NAc, nucleus accumbens; AMG, amygdala; HIP, hippocampus; HYPO, hypothalamus.

intractable. The challenge for any organism is to develop a ‘good-enough’ representation of self-in-world by selecting those aspects and characteristics of self-in-world that best facilitate adaptive behavior— that is, have value— and distributing the processing of those aspects to different regions for rapid, parallel processing. However, given that there is only one motor plant, these parallel models or representations must somehow converge and be integrated in decision-making that determines unitary organismal behavior— an organism

cannot go right and left at the same time. While we typically think of the problem of value as ascertaining what stimuli, events and actions yield something of utility to the organism (e.g., food, water, shelter, a mate), at a deeper level obtaining real-world value is contingent upon the adequacy of internal representations, that is, how well an organism models self-in-world. In this sense, the challenge of determining which stimuli, events and actions are valuable is preceded and predicated upon a more basic question: what representations— what *neural activity*— has value, that is, generates the most productive interactions with the environment?

The problem of coherence in distributed processing

A fundamental question in neuroscience is how such distributed, parallel partial models are integrated into a functional 'whole-brain' model to generate unitary organismal action (Breakspear, 2017; Christophel et al., 2017). This is not to suggest that at all times all neural substrates are operating perfectly together. There may be times when the activity of different brain regions are poorly coordinated, in conflict or even entirely unrelated. More generally, however, the problem is ensuring that distributed processing and representations of different aspects of self-in-world work together functionally to collectively maintain an adequate, composite model of both self and world. As corollary, coherence across neural substrates would necessitate that the learning and plasticity that shapes these partial, distributed models should *also* operate under some mechanism that promotes coherence within the composite, 'whole-brain' model.

There are numerous possible solutions to the problem of coherence in distributed processing. Different neural substrates tend to be densely interconnected such that fundamentally the activity of any given region is contingent upon input activity from other regions such that no neural substrate is an island, as it were. Hierarchical control relationships can also facilitate coherence. This sort of coherence-inducing responsibility might be ascribed to the prefrontal cortex (PFC), a locus of executive function and cognitive control (Fuster, 2001; Koechlin, 2016; Miller, 2000). While the PFC clearly plays a crucial role in coordinating both neural activity and behavior, the PFC is not a homunculus where all the processing across the brain converges to form an uber-model that, in turn, directs all neural activity (Hazy et al., 2007). Many regions contribute to decision-making and generating behavioral responses, including the hippocampus (Burton et al., 2009; Johnson et al., 2007; Yu and Frank, 2015), amygdala (Janak and Tye, 2015; Kim et al., 2017; Rudebeck et al.,

2017), parietal cortex (Cui, 2014; Platt and Glimcher, 1999), as well as many subcortical structures, including the basal ganglia (Bogacz and Gurney, 2007; Hikosaka et al., 2014) and midbrain regions. Further below, we briefly revisit the role of the PFC in the context of the hypothesis being developed here.

To facilitate coherence, we propose two functionally distinct integrative-convergence hubs contributing in two distinct ways. First, we suggest the basal ganglia mediate *discriminative selection*, where activity from different substrates (different models) is selectively modulated *in the context* of activities from other neural substrates (other models). Second, we suggest the midbrain serves as a consensus signal indicating coherence and 'agreement' between distributed models across afferent brain regions. This consensus signal might be likened to the conductor's baton: it keeps the various sections coordinated to produce a symphony, and with the wave of the baton, everyone knows what to do.

Midbrain - striatum as central axis of distributed coherence

Both the midbrain and the basal ganglia (via the striatum) are sites of convergence of broadly distributed input and both, in turn, modulate broadly distributed targets across the brain. Because of these characteristics, together with dense interconnectivity between the basal ganglia and the midbrain (Guo et al., 2015; Watabe-Uchida et al., 2012), we propose these substrates serve as a central axis of integration, with the basal ganglia and the midbrain serving the discriminative selection and consensus signaling integrative functions, respectively.

As our focus here is dopamine, we will make our case for the basal ganglia briefly. The notion that the striatum is an integrative substrate has been proposed by many for decades (Hikosaka et al., 2014; Kemp and Powell, 1971). Three characteristics of the basal ganglia make it suited for a discriminative selection mode of integration. First, it is a site of convergence, reflected in several related ideas. The convergence of afferent inputs has been proposed to serve a dimension reduction function (Bar-Gad et al., 2003). For example, a single MSN receives ~ 5000 cortical afferents (Alexander, 1994), including the convergence of two streams of input, corticostriatal and thalamostriatal, though individual cortical axons will make only 1-2 contacts with a single MSN (Tepper et al., 2007), suggesting tremendous convergence. Conversely, a single cortical axon can innervate up to 14% of the striatum (Zheng and Wilson, 2002). One elegant characterization of this distributed

convergence is the ‘mosaic of broken mirrors’ that suggest cortical representations are fragmented and distributed across the striatum allowing different combinatorial units (Da Cunha et al., 2009). In addition to the distributed convergence of information arising from a single afferent source— the mosaic— the striatum receives convergent inputs from multiple afferent substrates, an idea elegantly expressed by Humphry and Prescott (Humphries and Prescott, 2010) as a ‘crossroads,’ where neural activities from different regions intersect (see also, Choi et al., 2017; Hintiryan et al., 2016); in our parlance, models arising from different neural substrates are integrated to select the activities from each incoming model that is most relevant in the context of other models. Second, unlike dopamine where neurons come to fire similarly with learning, recordings from medium-spiny neurons show the opposite, that with learning (for example, in a conditioning task), different MSNs develop differential patterns of phasic activity in response to discrete stimuli and task events (Day et al., 2006; German and Fields, 2007; Hikosaka et al., 1989a, 1989b, 1989c; Nicola et al., 2004; Tremblay et al., 1998), suggesting discriminative learning. Finally, the basal ganglia in turn modulate their sources of input, illustrated best by cortico-basal ganglia-cortical loops where cortical activity drives basal ganglia activity which subsequently modulates cortical activity (Lee et al., 2016). In short, the basal ganglia provide a substrate whereby activity in different neural substrates— different models— can be modulated in the context of other neural substrates and models. For example, cortical activity can be modulated ‘in light of’ hippocampal or amygdalar activity. Gerraty et al (Gerraty et al., 2018) demonstrate that learning to associate a visual cue with reward involved dynamic changes in the coupling between striatum and multiple brain regions, including frontal and visual cortices, and that these changes were correlated with learning rate (see also den Ouden et al., 2010; Horga et al., 2015; van Schouwenburg et al., 2010). While the entire basal ganglia circuit is necessary for this integrative function to modulate efferent targets, we will focus on the striatum as the primary substrate of integration and selection (though in figures, we will retain BGS, basal ganglia system). Additionally, we will not differentiate striatal territories, acknowledging differences across these circuits in input-output wiring as well as subtle differences in physiology, gene expression and microcircuitry; however, these territories are more alike than different and here we focus on what we assert is the core function of the striatum across its territories: integrative discriminative selection. A dominant theory of the basal ganglia is that they select actions, arbitrating between competing demands on the motor plant (Redgrave et al., 1999). In the

hypothesis developed here, we also suggest the basal ganglia implements a selection mechanism; however, rather than selecting between complete representations of competing action choices —action A or action B— our hypothesis would consider action representation *per se* to be cortical, which is then fragmented and distributed across the striatum (Da Cunha et al., 2009) where it is integrated with activity from other brain regions which, in turn, modulates cortical representations through re-entrant loops. The selection process is, in essence, determining which 'pieces parts' from different afferent inputs and regions go together productively, rather than selecting between complete representations of action choices.

A central requirement for this proposed striatal function is some mechanism by which selection is guided, a teaching signal. Dopamine has long been believed to provide this teaching signal (Schultz, 1998; Schultz et al., 1997; Wise, 2004), but this leads us back to our original question: how does the midbrain know? In the next section, we describe dopamine as an integrative hub generating a consensus signal and propose how these two hubs together— accounting for their dense interconnectivity— facilitate neural coherence that allows a *whole brain model* to emerge from multiple, parallel models distributed across different neural substrates.

Midbrain consensus signaling: emergent synchrony in dopamine cell activity as Bayesian model selector

In concert with the basal ganglia as a hub for integrative discrimination, we propose that the midbrain serves as a hub for integrative consensus where the consensus signal generated by midbrain dopamine serves both a teaching and activational function. In this view, the apparent association between dopamine and reward arises as an epiphenomenon of consensus; that is, distributed models across afferent inputs to the midbrain converge on recognition of a stimulus as rewarding or predictive of reward. To accomplish this, we propose that midbrain dopamine functions at the intersection of two axes (Fig 3): (i) a largely excitatory axis receiving distributed input broadly across the brain, essentially 'sampling' whole brain activity, which drives midbrain (particularly bursting) activity and (ii) an inhibitory axis derived from the striatum, which serves to gate midbrain activity according to striatal integrative, discriminative selection. We propose the core function of midbrain dopamine— signaling consensus— arises from the emergence of congruence in these two axes such

that the excitatory drive is complemented by selective disinhibition in the inhibitory drive, facilitating model selection, synchrony and generation of a temporally resolved phasic signal.

Returning to our vector representations, we posit a random distribution of inhibitory inputs from the basal ganglia across dopamine cells, similarly as we suggested for excitatory inputs, generating input vectors for individual dopamine cells that represent a sampling of both excitatory and inhibitory inputs, here drawn separately for clearer conceptualization (Fig 4). We posit that *each* dopamine cell serves as a Bayesian

Figure 3 *Two axes model of dopamine integrative consensus signaling*. Proposed model where midbrain dopamine integrates two primary axes of input, (1) a (dis)inhibitory axis arising from the basal ganglia (ventral pallidum, striosomes, accumbens) and (2) a largely excitatory axis arising from distributed afferents across the brain, reflecting both top-down (e.g., cortical inputs, amygdala, hippocampus, BNST) and bottom-up information processing (e.g., collicular, multiple brainstem afferents). For specific inputs, see Watabe-Uchida (2012), Lerner (2015), Beier (2015). We propose a general principle that the intersection and integration of these (dis)inhibitory and excitatory axes provides a primary selection mechanism for midbrain dopamine cells in generating dopamine cell output. Additional complexities, such as inhibitory input on the 'excitatory axis' (e.g., lateral habenula) or interactions with other neuromodulators, such as serotonin and norepinephrine, would require accounting for in further development and elaboration of the basic model proposed.

evaluator of a randomly sampled model of afferent activity computing the probability that ‘something important’ is occurring in distributed neural processing. As greater numbers of dopamine cells— greater numbers of randomly sampled models of distributed activity— converge on and signal ‘something important’, then synchrony increases and with it temporally resolved fluctuations in extracellular activity. We will parse this more concretely:

(1) The probability being determined is whether ‘something important’ (clarified further below) is occurring in distributed neural activity—the input space— from which individual dopamine cells are sampling.

(2) The key inputs are the excitatory drive (E) and inhibitory drive (I) onto an individual dopamine cell, which we might think of as *advocate* and *skeptic*, respectively.

(3) The **prior** is the excitatory drive, $P(E)$, with greater activity reflecting greater probability of something important occurring. Notably, this prior can be modulated by previous experience either via learning in afferent regions that contribute the input or via synaptic plasticity at dopamine cell synapses that strengthen/weaken particular connections. Thus, activity associated with a cue light may be strengthened such that it has a greater impact on dopamine cell activity, reflecting a greater prior belief that neural activity arising in response to the cue light is important.

(4) The **posterior**— the dopamine cell response/signal, $P(E|I)$ — is the probability that the excitatory drive represents important distributed activity occurring across distributed brain regions given the inhibitory input from the basal ganglia.

(5) The **likelihood**, $P(I|E)/P(I)$, is a function of the basal ganglia input to the dopamine cell, arising as excitatory drive on the basal ganglia $P(I|E)$ modulates tonic inhibitory tone $P(I)$. More precisely, the signal would be disinhibitory, i.e., $P(dI|E)/P(dI)$. Normally at tonic inhibitory levels, $P(dI)$ would be low, increasing the impact of disinhibition arising as a consequence of excitatory drive on the basal ganglia, i.e., $P(dI|E)$. For simplicity, we retain simply $P(I|E)$ and $P(I)$.

It is worth noting that while we present here apparent mathematical quantities— $P(E)$ etc.— these are very different from discrete quantities proposed in prior models, such as $V(t)$. Critically, in positing that excitatory drive onto dopamine cells represent a prior belief that the current, on-going activity is important does not constrain the representational content of that activity. That activity, as Tian et al (2016) demonstrated,

could *reflect* value, prediction, prediction errors or, more to the point, activity arising entirely independent of value as an abstract quantity, such as signals reflecting proprioceptive or motor activity, motivational state/choice, arousal and so on. That is, as ‘evidence in favor of prior belief’ the representational content of afferent input could be broad. Moreover, a more agnostic view on the representational content driving P(E) opens the door to encoding value on different timescales, such as might arise when different contexts are associated with greater reward availability, or greater opportunity (Beeler and Mourra, 2018; Hamid et al., 2015; Howe et al., 2013).

Thus, our view is that midbrain dopamine is constantly sampling neural activity monitoring for occasions when the input from both the advocate, P(E) and the skeptic, P(I|E)/P(I) agree that the current activity distributed across the brain represents an opportunity to be acted upon and learned about. When a large percentage of dopamine cells arrive at the same ‘conclusion,’ synchrony arises and a temporally resolved fluctuation in extracellular dopamine encodes opportunity.

Distributed cascading learning: midbrain dopamine, the last to learn

We propose a learning cascade where learning occurs, roughly sequentially but interleaved, along three axes (Fig 5). First, primary learning occurs in afferent regions contributing to midbrain dopamine. That is, the models constructed in the prefrontal cortex, amygdala, hippocampus and so on undergo learning and modification, which alters the input driving the midbrain. Thus, as these afferent regions learn about the predictive value of a cue-light, this learning will be reflected in the input to the midbrain. Second, by integrating multiple models from different brain regions, as discussed above, the basal ganglia learns, which both alters the inhibitory drive on the midbrain but also modulates primary learning across multiple brain regions, which indirectly modifies the excitatory drive onto the midbrain. Third, the midbrain itself learns by selectively strengthening and weakening particular synaptic inputs. While it is known that afferent synapses onto dopamine cells can undergo synaptic plasticity (Lüscher and Malenka, 2011), exhibiting both LTP and LTD, the mechanisms regulating such plasticity are poorly understood. We would propose that plasticity at dopamine cells is primarily regulated by the balance between inhibitory and excitatory inputs (Tan et al., 2010), with disinhibition facilitating activity-dependent LTP at excitatory inputs. In the terminology we use above, dopamine

cells select strong excitatory inputs to strengthen when both the advocate (excitatory drive, $P(E)$) and the skeptic (disinhibition of inhibitory drive, $P(I|E)/P(I)$) are in agreement, effectively increasing the prior in future iterations of the stimulus. As midbrain dopamine learns, this in turn modulates both primary learning and learning in the basal ganglia (which in turn modulates primary learning). Crucially, though, this last link in the cascade where learning in midbrain dopamine modifies learning elsewhere is titrated by the extent to which the population of dopamine cells 'come to agreement' and fire synchronously, that is, $P(E|I)$ is scaled by the degree of synchrony across dopamine cells.

Figure 4. *Illustration of midbrain dopamine as a Bayesian model selector.* Random distribution of vectorized inputs from an afferent input space, as illustrated in Fig 1, where each dopamine cell samples on-going activity, mapped onto a Bayesian construal. The excitatory axis (green) is assigned as the prior (the 'advocate') and the (dis)inhibitory axis (red) is assigned as the likelihood (the 'skeptic'). The posterior (blue) arises from the integration of these two axes (Fig 3) at both the level of individual dopamine cells (firing rate) and at a population level as summed posteriors (firing rates), in which synchrony determines the degree to which increases in firing rates in individual cells sum to produce a population-based phasic signal, which we construe as a consensus index, both consensus across dopamine cells as Bayesian units and consensus across the sampled input space, reflecting widespread afferent activity in response to current stimuli.

In this construal midbrain dopamine cells are, as it were, the last to learn. This makes sense from the perspective of our hypothesis that midbrain dopamine signals *consensus*, implicitly suggesting ‘the rest of the brain’ has to learn first; however, this goes contrary to substantial evidence suggesting that dopamine is critical in mediating such learning in the first place, especially in the basal ganglia. To resolve this chicken-egg conundrum, not unique to our hypothesis, we will develop the notion of bootstrapping, an idea implicit in many theories of dopamine: learning always proceeds from what is already known.

Bootstrapping: distributed learning, distributed teaching

When an animal encounters an unexpected reward, dopamine generates a signal. From older learning theories (e.g., Wise, 2004), this is construed as an *unconditioned* stimulus, suggesting that no learning is required, perhaps that the response is hardwired. However, we would argue that the so-called unconditioned stimulus is, indeed, learned and represented in models across distributed brain regions; that is, it is *known*. In subsequent temporal difference theories, the dopamine signal arises as a consequence of *surprise*; that is, it is the unpredicted-ness of the reward that generates a signal (Schultz et al., 1997). But even here, it should be noted that the established *known* value of the unexpected reward is indispensable. In both cases, the dopamine teaching signal— that can play a major role in learning across the brain— arises in response to a stimulus (e.g., food pellet) that is well-known: on which there is consensus ‘this is valuable.’ This consensus signal of value, then, facilitates learning about the stimuli, events and actions that were antecedent or co-terminous with the obtained value in order to be able to develop predictive models.

Whether one construes a dopamine teaching signal as a reinforcement signal or temporal difference prediction error, either way learning is not immediate but occurs gradually over time. This can be accounted for with a simple idea: the neural activities associated with stimuli, events and actions co-terminous or antecedent to the *known* positive outcome are ‘tagged’ —synaptic connections strengthened— in afferent regions by increased dopamine stimulated by known, obtained value (Kreitzer and Malenka, 2008; Lovinger, 2010). We would argue that *all* antecedent/co-terminous neural activity targeted by dopamine is tagged in an activity-dependent way. Then, across repeated experiences (trials), those neural activities consistently antecedent to the outcome are recurrently strengthened, effectively selecting predictors of the outcome, while those neural

Figure 5. *Cascading learning*. Learning occurs progressively interleaved at three levels: (i) in primary models of individual afferent substrates, both top-down and bottom-up (blue), (i) in a secondary, integrative selection model in the basal ganglia system (orange) and (iii) in the midbrain dopamine system itself (green). Arrows indicate how learning at different levels influences learning at other levels.

activities that do not consistently precede the outcome are either not strengthened or weakened. Within the cascading learning described above, we would suggest this progressive learning occurs simultaneously, interleaved across all layers of learning. As learning progresses, a phasic dopamine signal emerges in response to the appropriate antecedent stimulus (e.g., cue light). The emergence of this cue-light response in dopamine, we suggest, reflects this distributed, cascading learning that is bootstrapped from established learning where midbrain dopamine plays a crucial role in signaling the *already known*. In this view, dopamine cells *are* the last to learn. Specifically, clear, temporally resolved phasic signals emerging from midbrain dopamine in response to a cue light arise from emergent synchrony. Emergent synchrony, in turn, arises as the sampled model of ‘activity in the rest of the brain’ in individual dopamine cells converge across the population of dopamine cells on the same aspects of the total afferent input space. In short, dopamine cells collectively come to recognize that the cue-light has a substantial, organizing impact on on-going afferent

neural activity across neural substrates and signals this consensus. At this point, the cue light is *known*— not just to midbrain dopamine, but broadly across the brain— and its appearance activates a phasic dopamine signal that mobilizes and energizes responding and further learning.

At root, this notion of bootstrapping is broadly consistent with current theories of dopamine. In both reinforcement and temporal difference perspectives, signaling about well-known stimuli facilitates learning about less known stimuli. In the next sections, we will focus on two cases that appear to present problems for our consensus hypothesis: the *loss* of phasic dopamine observed with well-predicted, expected reward and the increase in dopamine signaling associated with novelty.

The loss or reduction of phasic dopamine in response to well predicted rewards

The idea that dopamine serves as a reinforcement teaching signal has a long history and substantial evidence (Wise, 2004), including more recent work showing that activation of phasic dopamine is sufficient for associative learning (Steinberg et al., 2013). In this view, dopamine signals *value obtained*. In landmark work, Schultz and colleagues demonstrated that when there is a predictor for the reward, learning causes a shift of phasic dopamine from the reward to the predictive cue; that is, phasic dopamine no longer responds to reward when it is fully predicted (Schultz et al., 1993). Building on Rescorla-Wagner learning theory, this gave rise to the temporal difference learning theory of dopamine where dopamine, rather than signaling value obtained, signals errors in prediction of value, or surprise (Schultz et al., 1997). While there is substantial evidence supporting the notion that dopamine signals an RPE, accumulating evidence suggest that dopamine signaling may not be so simple. A complete review of the evidence for dopamine as RPE vs. other types of signals is beyond the scope of the current hypothesis (see Berke, 2018; Watabe-Uchida et al., 2017). Here we will focus on the one aspect of prediction error signaling that poses problems for our consensus hypothesis: that the phasic dopamine response can disappear with well-predicted rewards. We will make two arguments: (1) that phasic dopamine does not, generally, actually disappear in response to rewards but is only modulated by mismatches between expected and obtained rewards and (2) that such surprise—or disregard— modulation of phasic dopamine reflect responses across distributed regions afferent to midbrain dopamine.

In the original studies that inspired the prediction error hypothesis, the data inarguably show that the dopamine response to predicted rewards completely disappears (i.e., studies of Schultz and colleagues). Rather than uncovering the core characteristic of phasic dopamine signaling, this ‘pure’ prediction error phenomenon may reflect midbrain dopamine activity in extreme, non-natural conditions *not* indicative of its function in more complex, naturalistic, less controlled circumstances. In many subsequent studies, the phasic dopamine response to reward does *not* disappear, though it may be modulated by recent reward history, consistent with prediction errors signaling but suggesting a mixed signal of value *modulated* by expectation. In different papers, the emphasis placed on the *value* vs. *error* components of the dopamine signal vary. For example, Hamid et al (2015) conclude dopamine is primarily signaling instantaneous value with a prediction error riding on top of this signal (see also Parker et al., 2016; Satoh et al., 2003) . Studies in the last 10 years are replete with examples of mixed signals where the response to reward (obtained value) never entirely disappears. We would argue we do not have to explain the loss of phasic dopamine to predicted reward because this not a general phenomenon but an overgeneralization from highly specific, artificial experimental conditions. Instead, we would argue that obtained value is typically always signaled— *reflecting consensus of known value across afferent brain regions*— though it may be modulated by expectations, incorporating error signaling—value adjustments— into value signaling.

Nonetheless, we feel we do have to explain, within the consensus hypothesis, both the disappearance of dopamine signal in well-predicted reward as well as expectation-based modulation. We will start with the disappearance of phasic dopamine in response to predicted primary reward. While temporal difference theories argue this arises from a subtraction of expected from received reward, we will suggest an alternative: learned sequences. In the case of Schultz’s monkeys, through thousands of repeated trials with little variation and essentially no distractors, we would argue that the monkeys— and the afferent regions contributing to dopamine— have learned a sequence of ‘cue->delay->reward’ which has become ‘chunked’ (Graybiel, 1998) into a single unit, both behaviorally and perceptually. In this view, the dopamine signal at the beginning of the sequence ‘stands for’ all that follows; that is, it is not the cue and reward that are known but the sequence itself that is known. Moreover, this sequence has become habitual and, we would argue, the brain has become ‘habituated’ to elements within the sequence. When the outcome does not occur as expected, a negative

prediction error is evoked. In line with Tian et al (2016), we would argue that midbrain dopamine is not detecting this mismatch, but that some/all the (habituated) afferent regions are detecting the mismatch and delivering the prediction error to the midbrain. Similarly, we would interpret modulation of dopamine value signaling that is consistent with prediction errors to also reflect the sensitivity of afferent regions to expectations, all based on learned sequencing of events.

Having said this, we acknowledge that our explanation is incomplete and that we have not provided a detailed alternative account of these phenomena. This is because we are not willing to reject the more traditional accounts entirely. In recent work, Eshel et al (2015) argue that GABAergic neurons in the midbrain provide a subtractive expectation signal. We find this compelling and not necessarily inconsistent with our consensus hypothesis. Though speculative and beyond the scope of the current hypothesis, it seems possible that GABA cells in the midbrain may *also* form a model of distributed neural activity that provides a subtractive substrate for comparing expectations to outcome. Outlining such a hypothesis, however, is beyond the scope of the current paper and limited by considerably less knowledge about the GABA system within the midbrain, including comprehensive knowledge of its afferents, projections, role in behavior and connectivity and relationship with dopamine. Here we argue only that data associated with RPE accounts of dopamine are not incompatible with the consensus hypothesis, neither classic Schultzean data nor more recent data such as Eshel et al (2015).

Encountering the unfamiliar

Another known feature of dopamine our consensus hypothesis has to account for is the dopamine response to novelty (Horvitz, 2000; Overton et al., 2014), including unfamiliar stimuli and environments, a challenge for other theories of dopamine as well. Novel sensory information entering the brain does not have a learned path of propagation determined through prior experience; instead, we assume the novel information generates widespread activity across regions and distributed models as the brain tries to process this new encounter based on prior experience and knowledge, with features of new stimuli partially activating memory of previous stimuli encountered as the brain uses its existing models to learn about the unfamiliar. We would expect this increased activity across distributed brain regions in response to the unfamiliar stimuli to drive the

excitatory axis and generate bursting and small dopamine transients, consistent with spontaneous bursts/transients intermingled with tonic activity observed both in electrophysiological and electrochemical recordings (Marinelli and McCutcheon, 2014; Owesson-White et al., 2012); however we would not expect *synchronous* firing phase-locked with stimuli but randomly distributed bursts/transients throughout exploration and examination of novel stimuli. That is, this increased activity does not generate a temporally resolved signal arising from synchronous dopamine activity but instead the bursts/transients are distributed across time, effectively increasing tonic dopamine facilitating increased activity, exploration and plasticity.

A distinction should be made between a familiar environment into which a single, novel object has been placed (as often tested in rodents) and placement into an unfamiliar environment where all the stimuli are unfamiliar. In an unfamiliar environment, we would expect intense neural activity as the animal explores and learns about the environment, as described above. With a single object, increased dopamine will arise whenever the animal's attention is turned to the object. If the animal's attention occurs in brief, intermittent episodes, the increased dopamine arising from, for example, a one second sniff of the new object will give the appearance of a temporally resolved signal, arising as the animal turns to and sniffs the object and diminishing when it turns away. We suggest this is not phasic dopamine arising from synchrony but an increase in dopamine activity distributed over time— tonic activity— in which the period of time is brief and episodic: *faux* phasic dopamine. However, if the animal spends an extended episode sniffing/examining the object (e.g., 10 seconds), this will yield randomly distributed transients and increased dopamine across time without providing meaningful temporal signal, similar to exploring an unfamiliar environment.

This increased dopamine facilitates exploration (Beeler et al., 2010; Humphries, 2012; Kayser et al., 2015) and learning (Schultz et al., 1997; Steinberg et al., 2013; Wise, 2004). As distributed processing across the brain comes to learn the object/environment is not really interesting, the neural activity driven by the no-longer interesting stimuli will decline (corresponding to the animal largely ignoring the object, habituation) and no longer drive dopamine. If, however, the novel object/environment is found to be associated with stimuli of *known* value (e.g., finding food), this will initiate the bootstrapping process; for example, if food is found under the novel object or if the novel environment includes a cue light that when illuminated is followed by a food

pellet. By engaging bootstrapping, the dopamine system facilitates transforming unfamiliar stimuli to learned stimuli.

In sum, we are arguing that phasic, temporally resolved dopamine signals arising from *synchronous* activity represents a consensus signal and always arises from *known* stimuli. The 'stamping-in' hypothesis (Wise, 2004) provides a bootstrapping mechanism that utilizes known stimuli to learn about *unknown*, antecedent and coterminous stimuli in order to elaborate better models of self-in-world. The prediction error hypothesis (Montague et al., 1996; Schultz et al., 1997) places emphasis on whether the known stimulus was *expected* or not, which is to say whether the predictive value of stimuli preceding the known stimulus—the sequence-- was *also* known. Both hypotheses capture essential aspects of a consensus signal, which uses what the animal already knows to learn about what it does not know and build a better model of opportunity in the world.

So, what *does* dopamine encode?

Temporal-difference computational theories of dopamine have been largely based on the idea that dopamine signals arise from a 'model-free' system (Montague et al., 1996). The values associated with a state (e.g., 'cue light on') are derived from experience in which the basis of that value is inaccessible, i.e., there is no model of how that value was derived, sometimes called a 'cached' value. As a consequence, dopamine provides a 'pure' value-related signal in a 'common currency' that is independent of distinct sensory features of experience; that is, two rewards of equal value should be substitutable and generate the same dopamine signal. However, it has become increasingly clear in recent years that dopamine does not provide a pure, model-free signal of value (Takahashi et al., 2017); rather, it appears that dopamine signaling may well be model-based (Daw, 2012; Daw et al., 2011; Langdon et al., 2018) such that not only value or changes in value, but the identity and characteristics of sensory information associated with that value can modulate the signal. In addition, value may be represented on different timescales (Beeler and Mourra, 2018). Moreover, evidence is accumulating that midbrain dopamine may not be restricted solely to value-related signals. As discussed above, phasic dopamine activity has been shown to correlate to motor movement (Howe and Dombek, 2016), arousal (Eban-Rothschild et al., 2016), motivational state (Sato et al., 2003; Syed et al., 2016) and timing

(Pasquereau and Turner, 2015; Soares et al., 2016). While all of these could be construed as related to value, such data increasingly strain the concept of value-related signaling and suggest that dopamine is generating a richer, more complex signal than can be captured in a discrete quantity such as value or value prediction errors. Reflecting on the apparent range and promiscuity of dopamine signaling, rather than try to force this diverse data into a single interpretation of what dopamine encodes, we suggest instead any hypothesis of dopamine should allow and account for multi-dimensionality. Our hypothesis suggests the following:

(1) *primacy of temporal encoding*. We momentarily set aside the question of *what* dopamine signals to suggest the crucial characteristic of dopamine signaling is *when*, not *what*. The effects of dopamine vary by region but in a general sense, dopamine mobilizes and energizes responding, as well as facilitating learning. As a consensus signal, phasic dopamine is effectively saying '*now*,' which serves to activate and coordinate diverse target regions to collectively respond rapidly and with vigor in order to seize an afforded opportunity. The crucial question-- *when* is '*now*'—can be answered with varying degrees of temporal specificity: from precise, such as a cue light indicating in 1 second a reward will be available, to vague, such as being in an environment with a higher rate of reward availability (Hamid et al., 2015; Howe et al., 2013). In our consensus theory, the precision of *when* is determined by the degree of synchrony among dopamine neurons, such that high synchronous activity leads to temporally resolved phasic signaling while asynchronous activity, even if bursting, leads to increased dopamine distributed across a period of time, i.e., increased tonic. In this view, each dopamine neuron is sampling afferent activity building a model of *when* as a readout of the probability that distributed activity across the brain is 'important' *now*, $P(E|I)$. As more dopamine cells converge on the same temporal model, this reflects consensus (i.e., 'index of indexes') and, in turn, mobilizes and invigorates responding *now*. This consensus arises as a result of cascading learning as described above. The net result is that phasic dopamine reflects *this is what we know, act now*. That timing is an integral part of dopamine signaling is an established idea, particularly with prediction error theories (Langdon et al., 2018; Pasquereau and Turner, 2015; Soares et al., 2016; Starkweather et al., 2017). Here we suggest temporal characteristics are more than a feature of dopamine signaling, but its fundamental, primary function: the conductor's baton.

(2) *Bayesian model of opportunity*. Above we argue that each dopamine cell is a Bayesian computational unit indexing afferent activity to generate a temporal model, and that collectively a population of

dopamine cells serve as a Bayesian model selector, generating synchrony and temporally resolved phasic signals when individual, sample-based dopamine models converge. This begs the question, temporal model of *what*? While the obvious answer is probability of *value*, this term may be too limiting. Value is a quantitative, economic term and must necessarily be computed or evaluated against some reference, i.e., it is not absolute. Food is less valuable to a sated animal. How is value interpreted in dopamine signaling related to motor activity? What is the 'value' in a sequence of motor movements? Building on the idea discussed above, we argue that dopamine is not signaling an valuation of stimuli or events *per se* in an economic sense, but rather signaling the value of current neural activity and representations: that is, that the activity occurring in distributed afferents has previously proven to be advantageous and should be acted upon; that is, dopamine signals an opportunity. Of course, 'opportunity' will more often than not correspond to events, actions and stimuli associated with value in the traditional sense, but the center of gravity in the term 'opportunity' is an impetus to *act* rather than a quantitative representation of an abstract value to be used in computations. As discussed above, value continues to be the ghost in the machine, distributed and difficult to localize, likely computed in multiple ways and shaping learning. Opportunity, in contrast, is much more concrete, reflecting the appearance of actionable neural activity.

(3) *Representational content embedded in dopamine signal is unconstrained.* As noted above, the notion that dopamine represents a pure value-related signal (model-free) is being displaced by the observation that dopamine signals are contaminated with 'content', i.e., identity and features of actual stimuli, so-called model-based representations of value (Langdon et al., 2018; Takahashi et al., 2017). We wish to push this argument one step further and argue that dopamine cells are collecting *evidence*—the midbrain is an evidentiary system— and that the representational nature (the 'content') of that evidence is unconstrained: basically, it can be anything and can reflect whatever representational content is being processed in afferent regions driving midbrain dopamine activity. Returning to our analogy of a dopamine cell as an index of afferent activity like an index of stocks, if one stock has a disproportionate effect on the index, say an automobile company, this does not mean that the index is 'signaling automobile' but if you look at what the index corresponds to in actual economic activity, it will correspond to increased car sales. Similarly, as a cue light, via learning, comes to increasingly organize neural activity and behavior, the temporal model of the probability

of opportunity signaled by dopamine will increasingly correspond to the cue light. But as the many studies reviewed above suggest, the representational content in dopamine signals can vary greatly, depending upon the nature of the task. While in one sense this *does* represent a model— temporal model of opportunity in our view— we would argue it does *not* represent a model in the sense intended by the term 'model-based' which implies some implementation of connected graph of states and associated values which can be searched to compute future expected value and inform decision-making. While such models surely exist in some afferent regions, we would argue the extent to which they appear in dopamine signaling is determined only by the extent to which they provide *evidence* for the probability of an opportunity *now*. Starkweather et al (2017) has demonstrated that dopamine may signal *hidden* states; that is, the brain has constructed states that do not correspond directly to particular stimuli and these states drive dopamine signals rather than discrete stimuli. The notion of hidden states is wholly consistent with the hypothesis proposed here, with a slightly different interpretation: a 'hidden state' driving dopamine activity is composite afferent activity (sampling) but where the afferent activity (the prior) is not dominated by clearly discernible signals such as a cue-light. Without a dominant driver in afferent activity, the representational content is an amalgam of neural activity for which the corresponding stimuli are harder to identify, giving rise to a "hidden" state.

Revisiting the PFC: the three Cs axis-- control, coherence and consensus

Returning to the role of the PFC as a seat of executive control, we would like to sketch a necessarily speculative model that creates an 'axis of agency' through the brain by mediating control (i.e., executive, cognitive control), coherence and consensus. The PFC is a key locus of cognitive control or executive function (Fuster, 2001; Miller, 2000); however, 'control' and 'coherence' are not interchangeable terms. Coherence arises when activity in different neural substrates are complementary and mutually conducive to generating an action in response to on-going stimuli. Control is the ability to sustain or shift coherent, whole-brain activity. The PFC is richly endowed with the capacity to profoundly influence neural activity in other regions (Ferenczi et al., 2016; Fuster and Bressler, 2012; Koechlin, 2016; Miller and Cohen, 2001). Densely interconnected across cortical and subcortical regions, a cardinal function of the PFC is to either sustain in working memory neural activity representing transient stimuli or to shift contents of working memory (Durstewitz et al., 2000; Fuster

and Bressler, 2012; Miller and Cohen, 2001). This provides a mechanism by which the brain is not slavishly responding to on-going stimuli, but can pick and choose stimuli to sustain as neural representations which then have influence on neural processing—broadly across distributed models in the brain— independent of or in conjunction with the on-going flow of sensation. This does not necessarily make the PFC independent of the coherence generating mechanisms, but does provide a locus or node that is itself a site of highly convergent information with broad reaching efferent influence. Here we posit that the striatum is a central substrate of coherence. The PFC can modulate striatal function (Goto and Grace, 2005)—not only through its projections to the ventral to medial striatum (Humphries and Prescott, 2010), but also through a signal broadcast widely throughout the striatum (Calzavara et al., 2007; Haber et al., 2006). In turn, the striatum reciprocally modulates the PFC (Frank and Claus, 2006; Hazy et al., 2007; Pasupathy and Miller, 2005), directly through re-entrant loops (Alexander et al., 1986; Haber, 2003; Haber and Knutson, 2010), but also via modulation of structures that, in turn, send afferent projections to the PFC, including other cortical regions (Choi et al., 2017). In this sense, the PFC is embedded within a larger neural system for generating coherence but simultaneously in a position to have disproportionate influence on the contents of that coherence, biasing and regulating on-going neural processing across neural substrates. The special position for the PFC is even more striking when considering dopamine: midbrain dopamine and the PFC have a strong, reciprocal connection (Carr et al., 1999; Carr and Sesack, 2000; Lammel et al., 2008), potentially construed as a 'direct line' between control and consensus substrates. This notion of an axis of control, coherence and consensus, which we construe as an axis of agency, is illustrated in Fig 6.

Limitations and tests of hypothesis

Limitations. The foregoing represents an initial formulation of a novel hypothesis of midbrain dopamine function. As such, it is not a complete, fully formed explication of what the midbrain does, and cannot answer all questions that might be posed. It is a starting point. In order to retain a tractable scope in developing this hypothesis, there are several important areas that have been intentionally ignored for now, which we note. First, we focus on dopamine but it is becoming increasingly evident that both dopamine and midbrain GABA projection and interneurons form a functional midbrain unit (Brown et al., 2012; van Zessen et al., 2012),

including the recent hypothesis that the dopamine prediction error arises as a subtractive computation between GABA and dopamine cells in the midbrain (Eshel et al., 2015). Second, we have glossed over the extent to which midbrain dopamine may consist of functionally separate subpopulations that signal independently, possibly through different 'channels' with distinct, segregated projection targets (Lammel et al., 2014; Morales and Margolis, 2017; Roeper, 2013; Sanchez-Catalan et al., 2014; Willuhn et al., 2012). Third, our hypothesis would naturally lead to the question of a potential role for synchronized oscillations at various frequencies

Figure 6 *Axis of agency*. We posit the prefrontal cortex, basal ganglia and midbrain system comprise an axis of integrative substrates that confer unity of function to distributed processing, with the PFC providing executive control, the BGS providing coherence and the midbrain signaling consensus, which facilitates action and learning. In the diagram, distributed processing in the top-down cortical and bottom-up subcortical substrates are represented as thick black lines. Gray arrows represent interconnectivity between these axes and the PFC and BGS. The intersection of excitatory and inhibitory inputs in the midbrain is highlighted with green/red intersection lines. Interconnectivity between these substrates within the proposed axis of agency is represented by thick, colored arrow (green, excitatory, red inhibitory, blue dopamine).

between the midbrain and other brain regions. Fujisawa et al (Fujisawa and Buzsáki, 2011) have demonstrated 4 Hz oscillatory activity in the VTA that coupled with prefrontal oscillations. The authors suggest that midbrain dopamine may play a role in synchronizing this oscillatory activity across brain regions, but hypothesize this arises from glutamate transmission in DA cells. Though oscillatory activity might easily play a role in the hypothesis proposed here, such oscillatory activity in the midbrain has been surprisingly little studied and not introduced into the current hypothesis. Fourth, this hypothesis focuses entirely on the volume transmission aspect of dopamine release, with no consideration of intrasynaptic dopamine signaling. Though some have posited a role for fast synaptic transmission in dopamine signaling (Grace et al., 2007), we have no evidence disentangling synaptic from extrasynaptic transmission in dopamine signaling, though all data characterizing dopamine release in target regions is, effectively, measuring extrasynaptic transmission. What contribution direct synaptic transmission might make to behavior remains essentially unknown and is not considered in this hypothesis. Finally, it has become evident that dopamine cells transmit signals via multiple transmitters (Trudeau et al., 2014), including glutamate, GABA and sonic hedgehog (Gonzalez-Reyes et al., 2012), not incorporated into our hypothesis.

Tests of hypothesis. The prediction error hypothesis makes very clear, simple predictions: that phasic dopamine signaling will correspond to discrepancies in expected and actual outcomes. Nonetheless, twenty years later we are still testing this hypothesis (Watabe-Uchida et al., 2017), which has proven to be a rich driver of experimental investigation (Eshel et al., 2015; Hamid et al., 2015; Langdon et al., 2018). Remarkably, the extent to which it is correct continues to be a subject of active debate (Berke, 2018). We would expect no less for our hypothesis.

Tests of our hypothesis are different from those of current theories as we focus more on mechanisms and the question of how dopamine signals are derived rather than specifying what those signals encode. Although this does lead to proposing a different function— signaling consensus— we explicitly state that what that signal encodes ('content') is variable, task-dependent and multi-faceted. The challenge is that current theories that posit that dopamine signals convey one type of information (e.g., prediction error) do not explain the conditions under which other types of dopamine signals are observed. Strictly speaking, a study that shows the dopamine signal carries value information (Hamid et al, 2015) falsifies the hypothesis that dopamine serves

as a prediction error. In our view, data over the last 10 years has accumulated showing that the dopamine signal is multifaceted and can encode different kinds of information; thus, what is needed is a theory that can encompass and explain this characteristic observed across studies. For this, we pivoted to focus on precisely what most theories ignore: *how* the dopamine signal is derived or generated.

(1) *Synchronous dopamine activity emerges as a consequence of learning and is essential to phasic dopamine signaling across all types of tasks and behavior.* Data already support this premise and it is indeed believed that dopamine cells must fire *en masse* to generate and broadcast a signal. Thus, with regards to our hypothesis, it is less a matter of proving or disproving the existence of dopamine cell synchrony and more a matter of assessing its role in mediating dopamine signals. While neither value or prediction error can account for the full range of dopamine signals observed (for example, in motor movements and non-reward contexts), we predict that synchrony will without exception correspond to temporally resolved phasic signals and that loss of synchrony will lead to loss of phasic signals. In our view, understanding the mechanisms and development of synchronous firing provides a better framework for experimental work than abstract concepts such as value or prediction errors. We provide some example experiments:

(a) *Context dependent changes in dopamine signaling.* The prediction error hypothesis has a hard time explaining why dopamine increases in reward-associated task context (Hamid et al, 2015, Howe et al, 2013). Our hypothesis predicts that this arises from an increase in asynchronous bursting activity in an environment in which reward potential is increased and that with a cue that predicts reward, synchrony in dopamine cells will emerge in response to the cue as the animal learns, essentially already demonstrated by Kim et al (2012). However, we would further predict that were the animal placed in the test box repeatedly without reward the asynchronous bursting activity would decline as the animal habituates and comes to realize the environment affords no opportunity. Of course, this could be interpreted conceptually as signaling the availability of value/reward, but we are arguing specifically that such interpretations should arise secondarily to understanding the *mechanisms* of signaling under such tasks and conditions.

(b) *Reversal task.* While the prediction error hypothesis predicts that surprise in response to changed contingencies will be signaled as positive and negative prediction errors, allowing the animal to learn the new contingencies, this focuses on the dopamine signals that arise following expected outcomes. The hypothesis is

less clear on what happens to dopamine signaling during the preceding predictive stimuli during reversal learning. Our hypothesis would suggest that updating and learning new contingencies would correspond to a period of decreased dopamine cell synchrony preceding the animal's decision point (e.g., a cue) followed by a re-emergence of synchrony as the animal learns the new contingencies.

(c) *Contingency degradation.* How dopamine signals change in contingency degradation has been little studied and poses problems for the prediction error hypothesis. Random delivery of rewards, as they cannot be predicted, would presumably generate non-habituating surprise signals, but how these positive prediction errors would cause a decrease in the value in reward predictors or rewarded actions (lever presses) is difficult to explain with current theories. Our hypothesis predicts an increase in asynchronous bursting activity through a session of contingency degradation and a reduction of synchrony in response to reward-predicting cues. This is because an unexpected pellet ('known value' in our parlance) generates a dopamine signal that reinforces activity occurring at the time of delivery, but because a stable predictor to these random events does not exist, this dopamine signal does not consistently reinforce any particular activity in afferent regions, effectively generating 'confusion,' which in turn diminishes synchrony arising in response to the cue because the cue, in effect, only explains a portion of reward delivery. Moreover, we predict the ratio of synchrony/asynchrony will vary with the proportion of predicted/random rewards delivered.

We are suggesting a fundamental shift: instead of designing experiments that ask what the dopamine signal encodes, we propose designing experiments that ask how the dopamine signal is generated—when and how it arises rather than the interpretive what it means-- and synchrony provides a sort of Archimedean point for examining this question.

(2) *Cue-related value and error signals arise in afferent regions prior to their appearance as phasic variations in dopamine.* The importance of synchrony in our hypothesis is that it reflects an integrative function. We propose above that 'dopamine is the last to learn,' such that when dopamine generates a phasic signal it is reflecting 'consensus' across afferent regions; that is, it is reflecting now established learning, essentially across the brain. The prediction error hypothesis suggests that dopamine error signals facilitate learning in target regions, which would suggest that value-related signals arise first in dopamine. Which is correct can be tested by simultaneously recording from multiple afferent regions across learning (e.g., a cued

reward task) and simply evaluating whether value and prediction error signals associated with the cue arise first in afferent regions or phasic dopamine signals. We further predict that the lagging phasic dopamine will correspond with emergent synchrony across dopamine cells—and that if you interfere with this synchrony, learning will be slowed—but not blocked—in afferent regions because the 'bootstrapping' process we described above will be compromised.

(3) No single afferent region encodes quantities needed to generate value-related dopamine signals.

The prediction error hypothesis, as essentially an equation, is agnostic as to how error signals are physiologically computed, though subsequent theories (reviewed above) have been advanced. The landmark paper describing this theory (Schultz et al, 1997) suggested only that the midbrain served as a comparator between expected and obtained reward without detailing how this was accomplished. We argue that when value-related signals arise in phasic dopamine, this reflects an integration of incoming afferents—consensus—rather than a distinct computation in the midbrain taking discrete quantities computed elsewhere as input, as prior theories suggest. Thus, we predict that shutting down any single excitatory afferent input (e.g., using optogenetics to inhibit terminals from specific afferents) will not significantly change phasic dopamine signals in a task. This would be analogous to extending the Tian et al (2016) study and silencing the signals they observed in various regions contributing afferent input to DA.

(3) Midbrain synaptic plasticity is critically regulated by the intersection of excitatory and inhibitory inputs. Our hypothesis proposes that learning in the VTA depends upon selective disinhibition of basal ganglia inhibitory afferents in conjunction with excitatory activity. Curiously, synaptic plasticity has not been investigated in dopamine cells to the same extent as, for example, in medium-spiny neurons in the striatum. Our hypothesis predicts that the degree to which an excitatory input undergoes LTP is dependent upon disinhibition of GABA inputs, essentially functioning as a plasticity gate. This mechanism can be tested in slice physiology (e.g., Tan 2010). Demonstrating that GABA inputs control modulation of synaptic strength at excitatory inputs would provide a pillar in support of our hypothesis. Beyond slice, our hypothesis would predict that impairing *selective* disinhibition in basal ganglia inputs to the VTA—both by an across the board inhibition or activation—would impair the emergence of dopamine cells synchrony, temporally resolved phasic dopamine and, in turn, learning. Current theories, including the reward prediction and incentive-salience hypotheses do

not actually address how midbrain dopamine cell plasticity arises nor how it contributes to dopamine signaling.

(4) *Testing our Bayesian construal.* We construe dopamine output signal as a Bayesian posterior, which can be measured either in individual dopamine cells as changes in firing rate or across a population of dopamine cells as synchronous bursting activity. While current theories focus on what that output signal 'means,' our hypothesis remains agnostic on this point and focuses on the terms that determine that signal. We assign the terms of Bayes theorem specifically to excitatory input (prior, $P(E)$) and basal ganglia inhibitory input, with basal inhibition, $P(I)$, and selective disinhibition, $P(E|I)$, comprising the likelihood term. Each of these components can be selectively measured and manipulated, for example through optogenetic and photometric tools, and the impact on the emergence of phasic dopamine signals under different learning tasks can be assessed. Using photometry to simultaneously measure excitatory and inhibitory terminal activity (from basal ganglia), our hypothesis predicts that in a task requiring an animal to learn to associate a cue with a reward, across learning excitatory terminal activity will increase in response to the cue concomitant with a decrease in inhibitory terminal activity. In contrast, during intercue intervals, we would predict no change or an increase in inhibitory activity while excitatory activity may remain high, at least in early stages, as the animal continues to explore the environment or simply behave, doing non-task activities. We predict that synchronous dopamine cell activity and an emerging temporally resolved phasic signal will correlate to this simultaneous change in excitatory and inhibitory activity, which we expect could be fit to the Bayesian equation as proposed to account for the resulting dopamine cell output. In contrast, most efforts mapping the prediction error hypothesis to anatomy have suggested that inhibitory input from the basal ganglia should *increase* as it represents expected reward, which cancels out the excitation of received reward. More fundamentally, if the basal ganglia are contributing expected reward, then blocking these inputs should render the midbrain incapable of computing prediction errors and effectively silence prediction errors. In our hypothesis, this basal ganglia inhibitory input represents a likelihood term. Blocking it would alter the Bayesian computation we propose, but unlike in the case of the prediction error hypothesis, ours would not predict a loss of phasic dopamine signaling, only a modification in their computation and emergence across learning. Concretely, we predict that disrupting this input and impairing the likelihood term would slow dopamine-related learning and make it less efficient but not block it. The same experimental approach could assess and account for dopamine signaling in non-reward

tasks, such as running the rotarod, circumstances when less mechanistic, more interpretive theories of 'what dopamine means' become unfortunately vague.

Conclusion

The hypothesis proposed here is not intended to discard prior work and theories on how dopamine encodes value-related signals; rather, it is an alternative account of how such signals arise. This consensus hypothesis can be construed as largely consistent with extant theories that describe dopamine as signaling value or value prediction errors. It also opens the door for mixed signals of both value and prediction errors, consistent with Tian et al (2016), as well as leaves a door open for dopamine signals that do not fit neatly into value-related theories, as discussed above (Dodson et al., 2016; Howe and Dombeck, 2016; Pasquereau and Turner, 2015). At first glance, the hypothesis proposed here might be thought of as non-normative; however, the fundamental computational problem remains the same: how to determine value to guide decisions. Our hypothesis redistributes the implementational solution. That is, the value-optimizing algorithm remains the same, but the teaching signal that drives that algorithm is more distributed. Just as dopamine signaling is used by TD algorithms as a mechanism to optimize selection in the basal ganglia, our hypothesis requires some mechanism by which selection occurs in the midbrain—some guiding signal serving the same function in the midbrain as dopamine serves in the basal ganglia. We have proposed this arises from the intersection of excitatory and (dis)inhibitory axes, the advocate (prior) and skeptic (likelihood) in our Bayesian construal. There is no reason this proposed midbrain selection process cannot be, as selection in the basal ganglia has been, tackled from a normative perspective.

As others do, we see the core computational problem as that of determining and predicting value, or opportunity, and learning the best responses to such opportunities (i.e., the value of actions). A central problem in theories on how the brain solves this computational problem is locating the core evaluative mechanism—that is, fundamentally, where in the brain is 'value' *decided*? Answering this question poses a risk of introducing a homunculus, some key brain region from which all value flows, falling into a linear processing model where 'this region does this' and *then* 'that region does that' and so on. In solving this core computational problem, we see value not originating from a single brain region, but rather more as a reverberating echo—the ghost in the

machine— where the core problem of selecting the most valuable *neural activity* is recapitulated in every region of the brain in different forms; like a fractal, the selection problem looks the same whether you zoom in on a single region or zoom out and look at multiple regions. In our view, each region is fundamentally solving the same overarching problem with different architecture, properties, inputs and outputs: the question is how they are bound and brought into coherence. In this, we posit that midbrain dopamine plays a crucial role: recognizing and signaling when coherence is achieved. Otherwise, how does the brain know it is in agreement and should act?

ACKNOWLEDGEMENTS

This work was supported by NIDA DA046058 (JB). We thank the following people for helpful feedback on earlier versions of the manuscript: Jon Horvitz, Andy Delamater, Carolyn Pytte and Jake Jordan. We thank Michael J. Frank and Saleem Nicola for helpful discussions. The authors declare no competing financial interests.

REFERENCES

- Agnati LF, Zoli M, Strömberg I, Fuxe K (1995) Intercellular communication in the brain: Wiring versus volume transmission. *Neuroscience* 69:711–726.
- Alexander GE (1994) Basal ganglia-thalamocortical circuits: their role in control of movements. *J Clin Neurophysiol* 11:420–431.
- Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu Rev Neurosci* 9:357–381.
- Arbuthnott GW, Wickens J (2007) Space, time and dopamine. *Trends Neurosci* 30:62–69.
- Bar-Gad I, Morris G, Bergman H (2003) Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Prog Neurobiol* 71:439–473.
- Barto AG (1995) Adaptive critic and the basal ganglia In: *Models of Information Processing in the Basal Ganglia* , pp215–232. Cambridge, MA: MIT Press.
- Beeler JA, Daw N, Frazier CRM, Zhuang X (2010) Tonic Dopamine Modulates Exploitation of Reward Learning. *Front Behav Neurosci* 4.

- Beeler JA, Mourra D (2018) To Do or Not to Do: Dopamine, Affordability and the Economics of Opportunity. *Front Integr Neurosci* 12.
- Beier KT, Steinberg EE, DeLoach KE, Xie S, Miyamichi K, Schwarz L, Gao XJ, Kremer EJ, Malenka RC, Luo L (2015) Circuit Architecture of VTA Dopamine Neurons Revealed by Systematic Input-Output Mapping. *Cell* 162:622–634.
- Berke JD (2018) What does dopamine mean? *Nat Neurosci* 21:787–793.
- Berridge KC (2007) The debate over dopamine's role in reward: the case for incentive salience. *Psychopharmacology (Berl)* 191:391–431.
- Beyeler A, Chang C-J, Silvestre M, Lévêque C, Namburi P, Wildes CP, Tye KM (2018) Organization of Valence-Encoding and Projection-Defined Neurons in the Basolateral Amygdala. *Cell Rep* 22:905–918.
- Björklund A, Dunnett SB (2007) Dopamine neuron systems in the brain: an update. *Trends Neurosci* 30:194–202.
- Bogacz R, Gurney K (2007) The Basal Ganglia and Cortex Implement Optimal Decision Making Between Alternative Actions. *Neural Comput* 19:442–477.
- Bolam JP, Pissadaki EK (2012) Living on the edge with too many mouths to feed: Why dopamine neurons die. *Mov Disord* 27:1478–1483.
- Breakspear M (2017) Dynamic models of large-scale brain activity. *Nat Neurosci* 20:340–352.
- Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. *Neuron* 68:815–834.
- Brown J, Bullock D, Grossberg S (1999) How the Basal Ganglia Use Parallel Excitatory and Inhibitory Learning Pathways to Selectively Respond to Unexpected Rewarding Cues. *J Neurosci* 19:10502–10511.
- Brown MTC, Tan KR, O'Connor EC, Nikonenko I, Muller D, Lüscher C (2012) Ventral tegmental area GABA projections pause accumbal cholinergic interneurons to enhance associative learning. *Nature* 492:452–456.
- Buneo CA, Andersen RA (2006) The posterior parietal cortex: Sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia* 44:2594–2606.
- Burton BG, Hok V, Save E, Poucet B (2009) Lesion of the ventral and intermediate hippocampus abolishes anticipatory activity in the medial prefrontal cortex of the rat. *Behav Brain Res* 199:222–234.
- Calzavara R, Mailly P, Haber SN (2007) Relationship between the corticostriatal terminals from areas 9 and 46, and those from area 8A, dorsal and rostral premotor cortex and area 24c: an anatomical substrate for cognition to action: The striatum of action planning. *Eur J Neurosci* 26:2005–2024.
- Carr DB, O'Donnell P, Card JP, Sesack SR (1999) Dopamine Terminals in the Rat Prefrontal Cortex Synapse on Pyramidal Cells that Project to the Nucleus Accumbens. *J Neurosci* 19:11049–11060.

- Carr DB, Sesack SR (2000) Projections from the Rat Prefrontal Cortex to the Ventral Tegmental Area: Target Specificity in the Synaptic Associations with Mesoaccumbens and Mesocortical Neurons. *J Neurosci* 20:3864–3873.
- Choi EY, Tanimura Y, Vage PR, Yates EH, Haber SN (2017) Convergence of prefrontal and parietal anatomical projections in a connectional hub in the striatum. *NeuroImage* 146:821–832.
- Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D (2017) The Distributed Nature of Working Memory. *Trends Cogn Sci* 21:111–124.
- Cisek P (2007) Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos Trans R Soc B Biol Sci* 362:1585–1599.
- Cisek P, Kalaska JF (2010) Neural Mechanisms for Interacting with a World Full of Action Choices. *Annu Rev Neurosci* 33:269–298.
- Contreras-Vidal JL, Schultz W (1999) A predictive reinforcement model of dopamine neurons for learning approach behavior. *J Comp Neurosci* 6:191–214.
- Cragg SJ, Rice ME (2004) DANCING past the DAT at a DA synapse. *Trends Neurosci* 27:270–277.
- Cui H (2014) From Intention to Action: Hierarchical Sensorimotor Transformation in the Posterior Parietal Cortex. *eNeuro* 1.
- Da Cunha C, Wietzikoski EC, Dombrowski P, Bortolanza M, Santos LM, Boschen SL, Miyoshi E (2009) Learning processing in the basal ganglia: A mosaic of broken mirrors. *Behav Brain Res* 199:157–170.
- da Silva JA, Tecuapetla F, Paixão V, Costa RM (2018) Dopamine neuron activity before action initiation gates and invigorates future movements. *Nature* 554:244–248.
- Daw ND (2012) Model-based reinforcement learning as cognitive search: neurocomputational theories. *Cogn Search Evol Algorithms Brain* 195–208.
- Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron* 69:1204–1215.
- Day JJ, Wheeler RA, Roitman MF, Carelli RM (2006) Nucleus accumbens neurons encode Pavlovian approach behaviors: evidence from an autoshaping paradigm. *Eur J Neurosci* 23:1341–1351.
- den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal Prediction Error Modulates Cortical Coupling. *J Neurosci* 30:3210–3219.
- Dodson PD, Dreyer JK, Jennings KA, Syed ECJ, Wade-Martins R, Cragg SJ, Bolam JP, Magill PJ (2016) Representation of spontaneous movement by dopaminergic neurons is cell-type selective and disrupted in parkinsonism. *Proc Natl Acad Sci* 113:E2180–E2188.

- Dreyer JK (2014) Three Mechanisms by which Striatal Denervation Causes Breakdown of Dopamine Signaling. *J Neurosci* 34:12444–12456.
- Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD (2010) Influence of Phasic and Tonic Dopamine Release on Receptor Activation. *J Neurosci* 30:14273–14283.
- Dreyer JK, Hounsgaard J (2013) Mathematical model of dopamine autoreceptors and uptake inhibitors and their influence on tonic and phasic dopamine signaling. *J Neurophysiol* 109:171–182.
- Dreyer JK, Vander Weele CM, Lovic V, Aragona BJ (2016) Functionally Distinct Dopamine Signals in Nucleus Accumbens Core and Shell in the Freely Moving Rat. *J Neurosci* 36:98–112.
- Durstewitz D, Seamans JK, Sejnowski TJ (2000) Neurocomputational models of working memory 3:8.
- Eban-Rothschild A, Rothschild G, Giardino WJ, Jones JR, de Lecea L (2016) VTA dopaminergic neurons regulate ethologically relevant sleep–wake behaviors. *Nat Neurosci* 19:1356–1366.
- Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, Uchida N (2015) Arithmetic and local circuitry underlying dopamine prediction errors. *Nature* 525:243–246.
- Eshel N, Tian J, Bukwich M, Uchida N (2016) Dopamine neurons share common response function for reward prediction error. *Nat Neurosci* 19:479–486.
- Ferenczi EA, Zalocusky KA, Liston C, Grosenick L, Warden MR, Amatya D, Katovich K, Mehta H, Patenaude B, Ramakrishnan C (2016) Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science* 351:aac9698.
- Foster DJ, Morris RGM, Dayan P (2000) A model of hippocampally dependent navigation, using the temporal difference learning rule. *Hippocampus* 10:1–16.
- Frank MJ, Claus ED (2006) Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev* 113:300–326.
- Fujisawa S, Buzsáki G (2011) A 4 Hz Oscillation Adaptively Synchronizes Prefrontal, VTA, and Hippocampal Activities. *Neuron* 72:153–165.
- Fuster JM (2001) The Prefrontal Cortex—An Update: Time Is of the Essence. *Neuron* 30:319–333.
- Fuster JM, Bressler SL (2012) Cognit activation: a mechanism enabling temporal integration in working memory. *Trends Cogn Sci* 16:207–218.
- Garris PA, Ciolkowski EL, Pastore P, Wightman RM (1994) Efflux of Dopamine from the Synaptic Cleft in the Nucleus Accumbens of the Rat Brain. *J Neurosci* 14:6084–6093.
- Gauthier JL, Tank DW (2017) Context-invariant encoding of reward location in a distinct hippocampal population.

- German PW, Fields HL (2007) Rat Nucleus Accumbens Neurons Persistently Encode Locations Associated With Morphine Reward. *J Neurophysiol* 97:2094–2106.
- Gerraty RT, Davidow JY, Foerde K, Galvan A, Bassett DS, Shohamy D (2018) Dynamic flexibility in striatal-cortical circuits supports reinforcement learning. *J Neurosci* 2084–17.
- Gonon F (1997) Prolonged and extrasynaptic excitatory action of dopamine mediated by D1 receptors in the rat striatum in vivo. *J Neurosci* 17:5972–5978.
- Gonon F, Burie JB, Jaber M, Benoit-Marand M, Dumartin B, Bloch B (2000) Geometry and kinetics of dopaminergic transmission in the rat striatum and in mice lacking the dopamine transporter. *Prog Brain Res* 125:291–302.
- Gonzalez-Reyes LE, Verbitsky M, Blesa J, Jackson-Lewis V, Paredes D, Tillack K, Phani S, Kramer ER, Przedborski S, Kottmann AH (2012) Sonic Hedgehog Maintains Cellular and Neurochemical Homeostasis in the Adult Nigrostriatal Circuit. *Neuron* 75:306–319.
- Goto Y, Grace AA (2005) Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci* 8:805–812.
- Grace AA, Bunney BS (1983) Intracellular and extracellular electrophysiology of nigral dopaminergic neurons--3. Evidence for electrotonic coupling. *Neuroscience* 10:333–348.
- Grace AA, Floresco SB, Goto Y, Lodge DJ (2007) Regulation of firing of dopaminergic neurons and control of goal-directed behaviors. *Trends Neurosci* 30:220–227.
- Graybiel AM (1998) The basal ganglia and chunking of action repertoires. *Neurobiol Learn Mem* 70:119–136.
- Guo Q, Wang D, He X, Feng Q, Lin R, Xu F, Fu L, Luo M (2015) Whole-Brain Mapping of Inputs to Projection Neurons and Cholinergic Interneurons in the Dorsal Striatum. *PLOS ONE* 10:e0123381.
- Haber SN (2003) The primate basal ganglia: parallel and integrative networks. *J Chem Neuroanat* 26:317–330.
- Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci* 20:2369–2382.
- Haber SN, Kim K-S, Maily P, Calzavara R (2006) Reward-Related Cortical Inputs Define a Large Striatal Region in Primates That Interface with Associative Cortical Connections, Providing a Substrate for Incentive-Based Learning. *J Neurosci* 26:8368–8376.
- Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology* 35:4–26.
- Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD (2015) Mesolimbic dopamine signals the value of work. *Nat Neurosci* 19:117–126.

- Hart AS, Rutledge RB, Glimcher PW, Phillips PEM (2014) Phasic Dopamine Release in the Rat Nucleus Accumbens Symmetrically Encodes a Reward Prediction Error Term. *J Neurosci* 34:698–704.
- Haruno M, Kawato M (2006) Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning. *Neural Netw* 19:1242–1254.
- Hazy TE, Frank MJ, O'Reilly RC (2010) Neural mechanisms of acquired phasic dopamine responses in learning. *Neurosci Biobehav Rev* 34:701–720.
- Hazy TE, Frank MJ, O'Reilly RC (2007) Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system. *Philos Trans R Soc B Biol Sci* 362:1601–1613.
- Hikosaka O, Kim HF, Yasuda M, Yamamoto S (2014) Basal Ganglia Circuits for Reward Value–Guided Behavior. *Annu Rev Neurosci* 37:289–306.
- Hikosaka O, Sakamoto M, Usui S (1989a) Functional properties of monkey caudate neurons. I. Activities related to saccadic eye movements. *J Neurophysiol* 61:780–798.
- Hikosaka O, Sakamoto M, Usui S (1989b) Functional properties of monkey caudate neurons. II. Visual and auditory responses. *J Neurophysiol* 61:799–813.
- Hikosaka O, Sakamoto M, Usui S (1989c) Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J Neurophysiol* 61:814–832.
- Hintiryan H, Foster NN, Bowman I, Bay M, Song MY, Gou L, Yamashita S, Bienkowski MS, Zingg B, Zhu M, Yang XW, Shih JC, Toga AW, Dong H-W (2016) The mouse cortico-striatal projectome. *Nat Neurosci* 19:1100–1114.
- Horga G, Maia TV, Marsh R, Hao X, Xu D, Duan Y, Tau GZ, Graniello B, Wang Z, Kangarlou A, Martinez D, Packard MG, Peterson BS (2015) Changes in corticostriatal connectivity during reinforcement learning in humans: Corticostriatal Connectivity During Learning. *Hum Brain Mapp* 36:793–803.
- Horvitz JC (2000) Mesolimbocortical and nigrostriatal dopamine responses to salient non- reward events. *Neuroscience* 96:651–656.
- Houk J C, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement In: *Models of Information Processing in the Basal Ganglia* , pp249–270. Cambridge, MA: MIT Press.
- Howe MW, Dombeck DA (2016) Rapid signalling in distinct dopaminergic axons during locomotion and reward. *Nature* 535:505–510.
- Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM (2013) Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *Nature* 500:575–579.

- Humphries M (2012) Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. *Front Neurosci* 6.
- Humphries MD, Prescott TJ (2010) The ventral basal ganglia, a selection mechanism at the crossroads of space, strategy, and reward. *Prog Neurobiol* 90:385–417.
- Hyland BI, Reynolds JNJ, Hay J, Perk CG, Miller R (2002) Hyland et al. - 2002 - Firing modes of midbrain dopamine cells in the fre.pdf. *Neuroscience* 114:475–492.
- Janak PH, Tye KM (2015) From circuits to behaviour in the amygdala. *Nature* 517:284–292.
- Joel D, Niv Y, Ruppin E (2002) Actor–critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw* 15:535–547.
- Johnson A, van der Meer MA, Redish AD (2007) Integrating hippocampus and striatum in decision-making. *Curr Opin Neurobiol* 17:692–697.
- Joshua M, Adler A, Prut Y, Vaadia E, Wickens JR, Bergman H (2009) Synchronization of Midbrain Dopaminergic Neurons Is Enhanced by Rewarding Events. *Neuron* 62:695–704.
- Kakade S, Dayan P (2002) Dopamine: generalization and bonuses. *Neural Netw* 15:549–559.
- Kawato M, Samejima K (2007) Efficient reinforcement learning: computational theories, neuroscience and robotics. *Curr Opin Neurobiol* 17:205–212.
- Kayser AS, Mitchell JM, Weinstein D, Frank MJ (2015) Dopamine, locus of control, and the exploration-exploitation tradeoff. *Neuropsychopharmacology* 40:454–462.
- Kemp JM, Powell TP (1971) The connexions of the striatum and globus pallidus: synthesis and speculation. *Trans R Soc Lond B Biol Sci* 262:441–457.
- Kim J, Zhang X, Muralidhar S, LeBlanc SA, Tonegawa S (2017) Basolateral to Central Amygdala Neural Circuits for Appetitive Behaviors. *Neuron* 93:1464–1479.e5.
- Kim Y, Wood J, Moghaddam B (2012) Coordinated Activity of Ventral Tegmental Neurons Adapts to Appetitive and Aversive Learning. *PLoS ONE* 7:e29766.
- Kishida KT, Saez I, Lohrenz T, Witcher MR, Laxton AW, Tatter SB, White JP, Ellis TL, Phillips PEM, Montague PR (2016) Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proc Natl Acad Sci* 113:200–205.
- Ko D, Wanat MJ (2016) Phasic Dopamine Transmission Reflects Initiation Vigor and Exerted Effort in an Action- and Region-Specific Manner. *J Neurosci* 36:2202–2211.
- Koechlin E (2016) Prefrontal executive function and adaptive behavior in complex environments. *Curr Opin Neurobiol* 37:1–6.

- Komendantov AO, Canavier CC (2002) Electrical Coupling Between Model Midbrain Dopamine Neurons: Effects on Firing Pattern and Synchrony. *J Neurophysiol* 87:1526–1541.
- Kreitzer AC, Malenka RC (2008) Striatal Plasticity and Basal Ganglia Circuit Function. *Neuron* 60:543–554.
- Lammel S, Hetzel A, Häckel O, Jones I, Liss B, Roeper J (2008) Unique Properties of Mesoprefrontal Neurons within a Dual Mesocorticolimbic Dopamine System. *Neuron* 57:760–773.
- Lammel S, Lim BK, Malenka RC (2014) Reward and aversion in a heterogeneous midbrain dopamine system. *Neuropharmacology* 76:351–359.
- Langdon AJ, Sharpe MJ, Schoenbaum G, Niv Y (2018) Model-based predictions for dopamine. *Curr Opin Neurobiol* 49:1–7.
- Lee HJ, Weitz AJ, Bernal-Casas D, Duffy BA, Choy M, Kravitz AV, Kreitzer AC, Lee JH (2016) Activation of Direct and Indirect Pathway Medium Spiny Neurons Drives Distinct Brain-wide Responses. *Neuron* 91:412–424.
- Lerner TN, Shilyansky C, Davidson TJ, Evans KE, Beier KT, Zalocusky KA, Crow AK, Malenka RC, Luo L, Tomer R, Deisseroth K (2015) Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* 162:635–647.
- Li W, Doyon WM, Dani JA (2011) Acute in vivo nicotine administration enhances synchrony among dopamine neurons. *Biochem Pharmacol* 82:977–983.
- Lovinger DM (2010) Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. *Neuropharmacology* 58:951–961.
- Lüscher C, Malenka RC (2011) Drug-Evoked Synaptic Plasticity in Addiction: From Molecular Changes to Circuit Remodeling. *Neuron* 69:650–663.
- Mamad O, Stumpp L, McNamara HM, Ramakrishnan C, Deisseroth K, Reilly RB, Tsanov M (2017) Place field assembly distribution encodes preferred locations. *PLOS Biol* 15:e2002365.
- Marinelli M, McCutcheon JE (2014) Heterogeneity of dopamine neuron activity across traits and states. *Neuroscience* 282:176–197.
- Matsuda W, Furuta T, Nakamura KC, Hioki H, Fujiyama F, Arai R, Kaneko T (2009) Single Nigrostriatal Dopaminergic Neurons Form Widely Spread and Highly Dense Axonal Arborizations in the Neostriatum. *J Neurosci* 29:444–453.
- Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* 459:837–841.
- Miller EK (2000) The prefrontal cortex and cognitive control. *Nat Rev Neurosci* 1:59–65.
- Miller EK, Cohen JD (2001) An Integrative Theory of Prefrontal Cortex Function. *Annu Rev Neurosci* 24:167–202.

- Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J Neurosci* 16:1936–1947.
- Morales M, Margolis EB (2017) Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nat Rev Neurosci*.
- Morris G, Arkadir D, Alon Nevet, Eilon Vaadia, Hagai Bergman (2004) Coincident but Distinct Messages of Midbrain Dopamine and Striatal Tonically Active Neurons. *Neuron* 43:133–143.
- Moss J, Bolam JP (2008) A Dopaminergic Axon Lattice in the Striatum and Its Relationship with Cortical and Thalamic Terminals. *J Neurosci* 28:11221–11230.
- Nicholson C (1985) Diffusion from an injected volume of a substance in brain tissue with arbitrary volume fraction and tortuosity. *Brain Res* 333:325–329.
- Nicola SM, Yun IA, Wakabayashi KT, Fields HL (2004) Cue-Evoked Firing of Nucleus Accumbens Neurons Encodes Motivational Significance During a Discriminative Stimulus Task. *J Neurophysiol* 91:1840–1865.
- O'Neill P-K, Gore F, Salzman CD (2018) Basolateral amygdala circuitry in positive and negative valence. *Curr Opin Neurobiol* 49:175–183.
- Overton PG, Vautrelle N, Redgrave P (2014) Sensory regulation of dopaminergic cell activity: Phenomenology, circuitry and function. *Neuroscience* 282:1–12.
- Owesson-White CA, Roitman MF, Sombers LA, Belle AM, Keithley RB, Peele JL, Carelli RM, Wightman RM (2012) Sources contributing to the average extracellular concentration of dopamine in the nucleus accumbens: Extracellular dopamine concentration. *J Neurochem* 121:252–262.
- Parker NF, Cameron CM, Taliaferro JP, Lee J, Choi JY, Davidson TJ, Daw ND, Witten IB (2016) Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat Neurosci* 19:845–854.
- Pasquereau B, Turner RS (2015) Dopamine neurons encode errors in predicting movement trigger occurrence. *J Neurophysiol* 113:1110–1123.
- Pasupathy A, Miller EK (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433:873–876.
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238.
- Pryce CR (2018) Comparative evidence for the importance of the amygdala in regulating reward salience. *Curr Opin Behav Sci* 22:76–81.
- Redgrave P, Prescott TJ, Gurney K (1999) The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* 89:1009–1023.

- Roeper J (2013) Dissecting the diversity of midbrain dopamine neurons. *Trends Neurosci* 36:336–342.
- Roitman MF (2004) Dopamine Operates as a Subsecond Modulator of Food Seeking. *J Neurosci* 24:1265–1271.
- Rudebeck PH, Ripple JA, Mitz AR, Averbeck BB, Murray EA (2017) Amygdala Contributions to Stimulus–Reward Encoding in the Macaque Medial and Orbital Frontal Cortex during Learning. *J Neurosci* 37:2186–2202.
- Salamone JD, Correa M (2012) The Mysterious Motivational Functions of Mesolimbic Dopamine. *Neuron* 76:470–485.
- Sanchez-Catalan MJ, Kaufling J, Georges F, Veinante P, Barrot M (2014) The antero-posterior heterogeneity of the ventral tegmental area. *Neuroscience* 282:198–216.
- Satoh T, Nakai S, Sato T, Kimura M (2003) Correlated Coding of Motivation and Outcome of Decision by Dopamine Neurons. *J Neurosci* 23:9913–9923.
- Schultz W (2007) Behavioral dopamine signals. *Trends Neurosci* 30:203–210.
- Schultz W (1998) Predictive reward signal of dopamine neurons. *J Neurophysiol* 80:1–27.
- Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci* 13:900–913.
- Schultz W, Dayan P, Montague PR (1997) A Neural Substrate of Prediction and Reward. *Science* 275:1593–1599.
- Schultz W, Stauffer WR, Lak A (2017) The phasic dopamine signal maturing: from reward via behavioural activation to formal economic utility. *Curr Opin Neurobiol* 43:139–148.
- Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G (2017) Dopamine transients are sufficient and necessary for acquisition of model-based associations. *Nat Neurosci* 20:735–742.
- Soares S, Atallah BV, Paton JJ (2016) Midbrain dopamine neurons control judgment of time. *Science* 354:1273–1277.
- Stachenfeld KL, Botvinick MM, Gershman SJ (2017) The hippocampus as a predictive map. *Nat Neurosci* 20:1643–1653.
- Starkweather CK, Babayan BM, Uchida N, Gershman SJ (2017) Dopamine reward prediction errors reflect hidden-state inference across time. *Nat Neurosci* 20:581–589.
- Steinberg EE, Keiflin R, Boivin JR, Witten IB, Deisseroth K, Janak PH (2013) A causal link between prediction errors, dopamine neurons and learning. *Nat Neurosci* 16:966–973.
- Sulzer D, Cragg SJ, Rice ME (2016) Striatal dopamine neurotransmission: Regulation of release and uptake. *Basal Ganglia* 6:123–148.
- Syed ECJ, Grima LL, Magill PJ, Bogacz R, Brown P, Walton ME (2016) Action initiation shapes mesolimbic dopamine encoding of future rewards. *Nat Neurosci* 19:34–36.
- Takahashi YK, Batchelor HM, Liu B, Khanna A, Morales M, Schoenbaum G (2017) Dopamine Neurons Respond to Errors in the Prediction of Sensory Features of Expected Rewards. *Neuron* 95:1395–1405.e3.

- Takahashi YK, Langdon AJ, Niv Y, Schoenbaum G (2016) Temporal Specificity of Reward Prediction Errors Signaled by Putative Dopamine Neurons in Rat VTA Depends on Ventral Striatum. *Neuron* 91:182–193.
- Takiyama K (2015) Sensorimotor transformation via sparse coding. *Sci Rep* 5.
- Tan CO, Bullock D (2008) A Local Circuit Model of Learned Striatal and Dopamine Cell Responses under Probabilistic Schedules of Reward. *J Neurosci* 28:10062–10074.
- Tan KR, Brown M, Labouèbe G, Yvon C, Creton C, Fritschy J-M, Rudolph U, Lüscher C (2010) Neural bases for addictive properties of benzodiazepines. *Nature* 463:769–774.
- Tepper JM, Abercrombie ED, Bolam JP (2007) Basal ganglia macrocircuits In: *Progress in Brain Research* , pp3–7. Elsevier.
- Tian J, Huang R, Cohen JY, Osakada F, Kobak D, Machens CK, Callaway EM, Uchida N, Watabe-Uchida M (2016) Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. *Neuron* 91:1374–1389.
- Tremblay L, Hollerman JR, Schultz W (1998) Modifications of Reward Expectation-Related Neuronal Activity During Learning in Primate Striatum. *J Neurophysiol* 80:964–977.
- Trudeau L-E, Hnasko TS, Wallén-Mackenzie Å, Morales M, Rayport S, Sulzer D (2014) The multilingual nature of dopamine neurons In: *Progress in Brain Research* , pp141–164. Elsevier.
- van Schouwenburg MR, den Ouden HEM, Cools R (2010) The Human Basal Ganglia Modulate Frontal-Posterior Connectivity during Attention Shifting. *J Neurosci* 30:9910–9918.
- Vandecasteele M (2005) Electrical Synapses between Dopaminergic Neurons of the Substantia Nigra Pars Compacta. *J Neurosci* 25:291–298.
- van Zessen R, Phillips JL, Budygin EA, Stuber GD (2012) Activation of VTA GABA Neurons Disrupts Reward Consumption. *Neuron* 73:1184–1194.
- Venton BJ, Zhang H, Garris PA, Phillips PEM, Sulzer D, Wightman RM (2003) Real-time decoding of dopamine concentration changes in the caudate-putamen during tonic and phasic firing: Decoding dopamine neurotransmission. *J Neurochem* 87:1284–1295.
- Vitay J, Hamker FH (2014) Timing and expectation of reward: a neuro-computational model of the afferents to the ventral tegmental area. *Front Neurobotics* 8.
- Watabe-Uchida M, Eshel N, Uchida N (2017) Neural Circuitry of Reward Prediction Error. *Annu Rev Neurosci* 40:373–394.
- Watabe-Uchida M, Zhu L, Ogawa SK, Vamanrao A, Uchida N (2012) Whole-Brain Mapping of Direct Inputs to Midbrain Dopamine Neurons. *Neuron* 74:858–873.

- Whitlock JR (2014) Navigating actions through the rodent parietal cortex. *Front Hum Neurosci* 8.
- Willuhn I, Burgeno LM, Everitt BJ, Phillips PE (2012) Hierarchical recruitment of phasic dopamine signaling in the striatum during the progression of cocaine use. *Proc Natl Acad Sci* 109:20703–20708.
- Wilson CJ, Young SJ, Groves PM (1977) Statistical properties of neuronal spike trains in the substantia nigra: cell types and their interactions. *Brain Res* 136:243–60.
- Wise RA (2004) Dopamine, learning and motivation. *Nat Rev Neurosci* 5:483–494.
- Yetnikoff L, Lavezzi HN, Reichard RA, Zahm DS (2014) An update on the connections of the ventral mesencephalic dopaminergic complex. *Neuroscience* 282:23–48.
- Yu JY, Frank LM (2015) Hippocampal–cortical interaction in decision making. *Neurobiol Learn Mem* 117:34–41.
- Zheng T, Wilson CJ (2002) Corticostriatal Combinatorics: The Implications of Corticostriatal Axonal Arborizations. *J Neurophysiol* 87:1007–1017.
- Zoli M, Torri C, Ferrari R, Jansson A, Zini I, Fuxe K, Agnati LF (1998) The emergence of the volume transmission concept. *Brain Res Rev* 26:136–147.